(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
15 May 2003 (15.05.2003)

PCT

(10) International Publication Number
WO 03/040393 A2

(51) International Patent Classification[7]: C12Q

(21) International Application Number: PCT/IB02/04615

(22) International Filing Date:
4 November 2002 (04.11.2002)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/332,633 6 November 2001 (06.11.2001) US

(71) Applicant *(for all designated States except US)*: **DECODE GENETICS EHF.** [IS/IS]; Sturlugotu 8, IS-101 Reykjavik (IS).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: **MARTINEZ, Roger, Alfonso, Moraga** [ES/DE]; Bleichstrasse 6, 60313 Frankfurt am Main (DE). **SIGURDSSON, Gunnar, Thor** [IS/IS]; Kjarrholmi 8, IS-200 Kopavogur (IS).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Published:
— *without international search report and to be republished upon receipt of that report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 03/040393 A2

(54) Title: NUCLEIC ACIDS ENCODING PROTEASES

(57) Abstract: Nucleic acids encoding proteases are disclosed, and methods of using same.

NUCLEIC ACIDS ENCODING PROTEASES

RELATED APPLICATION

This application claims priority to U.S Provisional Application 60/332,633, filed November 6, 2001, the entire teachings of which are incorporated herein by reference.

BACKGROUND OF THE INVENTION

A "protease" or "protease enzyme" is a protein which has proteolytic activity. These enzymes have important applications in industry, biology and medicine.

Industrial applications of proteases include food processing (*e.g.*, serine proteases are used to produce dairy products and protein-rich concentrates from fish and livestock), brewing, and alcohol production, and are common additives in laundry detergents (*e.g.*, the serine protease subtilisin is used to remove protein-based stains). They are also used in research, *e.g.*, to degrade undesirable proteins during purification processes.

In nature, proteases are involved in a wide array of biological pathways, in regulatory capacities, as treatments, or in pathological capacities. Tissue plasminogen activator (t-PA) and thrombin, for instance, are commonly used in pharmaceutical settings, while improper regulation of elastase can result in lung problems such as emphysema. Proteases have also been implicated in the occurance of many diseases, including Alzheimer's disease, cystic fibrosis, muscular dystrophy, arthritis, emphysema and other respiratory ailments, gastrointestinal diseases, hypertension, degenerative skin disorders, tumor invasion and metastasis and also viral-associated diseases, where they are important in viral maturation. The aspartic acid proteases pepsin and gastricsin are secreted into the stomach for food digestion, and are diagnostic indicators for stomach ulcer and cancer. Collagen and elastase degrade, respectively, the structural proteins collagen and elastin. Many

elastase degrade, respectively, the structural proteins collagen and elastin. Many human pathogens (*e.g.*, *Haemophilus influenza, Neisseria gonorrhoeae, Streptococcus sanguis, Streptococcus pneumonia, Neisseria meningitides*) possess proteases capable of cleaving the hinge region of human IgA1, which is a class of antibodies found in mucous. These proteases are therefore targets in preventing infections and pathogenesis by these organisms.

Proteases are also involved in cellular responses to starvation, heat-shock, and other stresses in which cells might find it advantageous to break down proteins to salvage component amino acids.

SUMMARY OF THE INVENTION

The present invention relates to human protease genes, particularly nucleic acids comprising protease genes, and the amino acids encoded by such nucleic acids. These sequences are shown in Appendix I. In Appendix I, each protease entry lists the University of California at Santa Cruz contig designation from which the sequence was analyzed (*e.g.*, "ctgchr11q_1"), the name (*e.g.*, "MOOSE13873"), the exon locations (*e.g.*, "7240429 . .7240464, . . ."), following by the amino acid sequence and the nucleic acid sequence.

Sub-family information on the sequences is shown in Appendix II. For each sequence, the following information is provided: the name (*e.g.*, "MOOSE13873"), the University of California at Santa Cruz contig designation from which the sequence was analyzed (*e.g.*, "ctgchr11q_1"), and the subfamily to which the sequence appears to belong. The assignments were made on the basis of the best E-value with which the sequence aligned.

In one embodiment, the isolated nucleic acid molecule comprises a nucleotide sequence selected from the group consisting of SEQ ID NOs:1-47, as shown in Appendix I, and the complements thereof. The invention further relates to a nucleic acid molecule which hybridizes under high stringency conditions to a nucleotide sequence selected from the group consisting of SEQ ID NOs:1-47, as shown in Appendix I, and the complements thereof. The invention additionally relates to isolated nucleic acid molecules (*e.g.*, cDNA molecules) encoding a

protease polypeptide (*e.g.*, encoding a polypeptide selected from the group consisting of SEQ ID NOs:48-94, as shown in Appendix I).

The invention further provides a method for assaying a sample for the presence of a nucleic acid molecule encoding all or a portion of a protease in a sample, comprising contacting said sample with a second nucleic acid molecule comprising a nucleotide sequence encoding a protease polypeptide (*e.g.*, one of SEQ ID NOs:1-47, as shown in Appendix I, or the complement of one of SEQ ID NOs:1-47; a nucleotide sequence encoding one of SEQ ID NOs:48-94, as shown in Appendix I), or a fragment or derivative thereof, under conditions appropriate for selective hybridization. The invention additionally provides a method for assaying a sample for the level of expression of a protease polypeptide, or fragment or derivative thereof, comprising detecting (directly or indirectly) the level of expression of the protease polypeptide, fragment or derivative thereof.

The invention also relates to a vector comprising an isolated nucleic acid molecule of the invention operatively linked to a regulatory sequence, as well as to a recombinant host cell comprising the vector. The invention also provides a method for preparing a polypeptide encoded by an isolated nucleic acid molecule described herein (a protease polypeptide), comprising culturing a recombinant host cell of the invention under conditions suitable for expression of said nucleic acid molecule.

The invention further provides an isolated polypeptide encoded by isolated nucleic acid molecules of the invention (*e.g.*, protease polypeptide), as well as fragments or derivatives thereof. In a particular embodiment, the polypeptide comprises an amino acid sequence selected from the group consisting of SEQ ID NOs:48-94, as shown in Appendix I. The invention also relates to an isolated polypeptide comprising an amino acid sequence which is greater than about 90 percent identical to an amino acid sequence selected from the group consisting of SEQ ID NOs:48-94, preferably about 95 percent identical.

The invention also relates to an antibody, or an antigen-binding fragment thereof, which selectively binds to a polypeptide of the invention, as well as to a method for assaying the presence of a polypeptide encoded by an isolated nucleic

acid molecule of the invention in a sample, comprising contacting said sample with an antibody which specifically binds to the encoded polypeptide.

The invention further relates to methods of diagnosing a predisposition to a condition mediated by a protease. The methods of diagnosing such a predisposition in an individual include detecting the presence of a mutation in a protease, as well as detecting alterations in expression of a protease polypeptide, such as the presence of different splicing variants of protease polypeptides. The alterations in expression can be quantitative, qualitative, or both quantitative and qualitative.

The invention additionally relates to an assay for identifying agents which alter (*e.g.*, enhance or inhibit) the activity or expression of one or more protease polypeptides. For example, a cell, cellular fraction, or solution containing a protease polypeptide or a fragment or derivative thereof, can be contacted with an agent to be tested, and the level of protease polypeptide expression or activity can be assessed. The activity or expression of more than one protease polypeptide can be assessed concurrently (*e.g.*, the cell, cellular fraction, or solution can contain more than one type of protease polypeptide, such as different splicing variants, and the levels of the different polypeptides or splicing variants can be assessed).

In another embodiment, the invention relates to assays to identify polypeptides which interact with one or more protease polypeptides. In a yeast two-hybrid system, for example, a first vector is used which includes a nucleic acid encoding a DNA binding domain and also a protease polypeptide, splicing variant, or fragment or derivative thereof, and a second vector is used which includes a nucleic acid encoding a transcription activation domain and also a nucleic acid encoding a polypeptide which potentially may interact with the protease polypeptide, splicing variant, or fragment or derivative thereof (*e.g.*, a protease polypeptide binding agent or receptor). Incubation of yeast containing both the first vector and the second vector under appropriate conditions allows identification of polypeptides which interact with the protease polypeptide or fragment or derivative thereof, and thus can be agents which alter the activity of expression of a protease polypeptide.

Agents that enhance or inhibit protease polypeptide expression or activity are also included in the current invention, as are methods of altering (enhancing or

inhibiting) protease polypeptide expression or activity by contacting a cell containing protease and/or polypeptide, or by contacting the protease polypeptide, with an agent that enhances or inhibits expression or activity of protease or polypeptide.

Additionally, the invention pertains to pharmaceutical compositions comprising the nucleic acids of the invention, the polypeptides of the invention, and/or the agents that alter activity of protease polypeptide. The invention further pertains to methods of treating conditions mediated by proteases, by administering protease therapeutic agents, such as nucleic acids of the invention, polypeptides of the invention, the agents that alter activity of protease polypeptide, or compositions comprising the nucleic acids, polypeptides, and/or the agents that alter activity of protease polypeptide.

DETAILED DESCRIPTION OF THE INVENTION

The present invention relates to nucleic acids comprising proteases, and the protease amino acids encoded by those nucleic acids.

Proteases are enzymes having proteolytic activity, and so comprise an extremely large class of enzymes having a wide array of functions. Outside of the cell, the proteolytic activity of proteases can be used in many different industrial processes, including food and beverage processing (*e.g.*, beer brewing, manufacture of dairy products, protein processing, etc.), and in detergent formulation (*e.g.*, for removal of proteinaceous stains). New versions of such proteases possessing increased levels of activity or binding, would therefore be of increased utility in such industrial settings. Likewise, those proteases with pharmaceutical application, *e.g.*, t-PA and thrombin, would also be more useful if versions were found possessing increased activity and/or more specific activity.

Proteases are also involved in many cellular functions, in both humans and other organisms, of which some of the functions can lead to disease conditions. Where proteases are associated with pathogenesis, it is therefore desirable to interfere with the activity of such enzymes. For instance, where a disease or undesirable condition is due at least in part to the action of or mediation by one or

more protease enzymes (*e.g.*, emphysema, cystic fibrosis, hypertension, etc.), a potential method of prophylaxis would be to prevent the protease associated with the disease from acting at all, or from acting at the time and/or with the ligand necessary to cause the disease or condition. For instance, for those pathogens which infect humans, *e.g.*, by cleaving the hinge region of human IgA1, pathogenesis could be prevented by inhibiting the action of that protease in the pathogenic organism.

With the availability of complete genomic sequences for many organisms today, including *Homo sapiens*, it has become clear that there is a need for data mining techniques to extract the information in them, *e.g.*, gene prediction programs. Of these, the most successful ones are those based on the comparison of known protein or protein-derived information, or those that use expressed sequence tags (ESTs) to predict gene location and structure.

One such algorithm is GeneWise. It bases its exon prediction on the use of Hidden Markov Models (HMMs) of proteins to be compared against a genomic sequence, so that the translation of the sequence will match the model in a similar way to other HMM profile searches (Eddy, *Curr. Opin. Struct. Biol.* 6(3):361-5 (1996), and allowing the presence of long insertions as long as they include donor and acceptor site sequences at both ends.

To take advantage of the algorithm, the models for different protein families must be built so that they represent the full-length sequences instead of the most common features in them. This is a major difference with existing HMM databases such as Pfam (Sonnhammer *et al.*, *Proteins* 28(3):405-20 (1997), in which each model is built to represent a family of proteins as broad as possible with minimum overlap between them.

In the present approach, the sequences were subdivided in several families so that the similarity inside of a group of them was over 50%. Given this approach, there are several points of overlap between different families when analyzing a sequence, so the discrimination must be done after the search is completed.

Several resources that include expert-supervised classifications are used to select the best groups of sequences, *e.g.*, the GPCRdb (Horn *et al.*, *Nucleic Acids Res.* 26(1):275-9 (1998)), PKR (Smith *et al.*, *Trends Biochem. Sci.* 22(11):444-6

(1997)), NuclearRdb (Horn *et al.*, *Nucleic Acids Res.* 29:346-349 (2001)), IOCH (Le Novere *et al.*, *Nucleic Acids Res.* 27(1):340-2 (1999)), Enzyme (Bairoch, *Nucleic Acids Res.* 28:304-305 (2000)) and Swiss-Prot (Bairoch *et al.*, *Nucleic Acids Res.* 28:45-48 (2000)). When none is available, or the sequences included in some groups are too distantly related, the grouping must be done manually, using the ClustalW (Thompson *et al.*, *Nucleic Acids Res.* 22:4673-4680 (1994)) package to measure the distance between different sequences.

The present model was built from multiple sequence alignments of the different protein families obtained with DiAlign 2 (Morgenstern, *Bioinformatics* 15(3):211-8(1999)). DiAlign works based on segment-to-segment comparisons instead of arbitrary thresholds for gap opening and extension, which makes it ideally suited for building models that represent an entire, full-length sequence, since the alignments built this way have more match states that would be assigned as insertion states when using other alignment algorithms. The models were built using the standard HMMer package.

To search for new genes, a genome-wide scan was done on the the University of California at Santa Cruz sequences, using the GeneWise algorithm. It translates the genomic sequence on the fly to proteins and can therefore maintain a reading frame through insertions and deletions. The algorithm also rewards gaps in the genomic sequence relative to the model if they are encapsulated within introns, like splice structure.

For each superfamily of proteins, a classification was obtained in which the sequences are grouped by length and similarity. Each one of these groups was then used to build a HMM profile representing this group of sequences. This approach aims to have models that can represent the full length of the encoded proteins for a whole range of proteins, without being too specific for any one of them or being too general, as would be a HMM built for large groups of sequences. This classification was based either on existing expert-supervised classifications, or by retrieval of sequences and classification based on pairwise alignment distances.

These models were then searched against the August 2001 fixed release of the Santa Cruz contigs using the Paracel GeneMatcher+ Hardware Accelerator with

the GeneWise algorithm. The sequences were chopped at 100 Kb with an overlap of 1 Kb. Each one of the superfamilies required between 3 and 6 days to complete and generate results. The results represent the coding regions of the complete final protein as it is found in the organism.

The cross-validation of the results was done in two steps. First, all of the hits with an E-value lower than 10^{-8} that do not overlap with one another were selected, and in the event of overlapping, the one with lowest E-value was selected. After selecting all of those matches, the DNA sequences were compared against the RefSeq database (Pruitt *et al.*, *Trends Genet.* 16(1):44-47 (2000)) using BLAST (Altschul *et al.*. *Nucleic Acids Res.* 25:3389-3402 (1997)).

Over 80% of the sequences were 90% or more identical to an existing human RefSeq entry. The differences are usually due to picking the wrong model for a certain sequence that appears as a hit more than once in different families, being a different valid splice variant, which can be tested by comparing to the EST database, or by addition of a small last exon to complete the match instead of accept an stop codon in a previous one. In all of such cases, the results are easily and quickly corrected by eye. Very rarely the algorithm will actually make a wrong prediction, which is consistent with the expected behaviour (Guigo *et al.*, *Genome Res.* 10(10):1631-42 (2000)).

Of the remaining sequences, over 50% have a match over 90% identical in the public domain protein databases, and the differences between those sequences in the databases and the potential translations is basically the same as the differences between the DNA sequences and the RefSeq entries.

A number of the genes were found to be linked with markers known to be associated with human diseases genes. These are shown in Appendix III. The diseases were linked to the HMM genes in the following manner: (1) the HMM gene models were compared to the consensus of the human genome sequence, located and the results kept in a relational database; (2) all possible markers (Sequence Tagged Sites (STS's)) (public or deCODE genetics) are also located in the same consensus using ePCR or BLAT and results kept in a relational database; and (3) LOD scores for diseases are linked to markers. A span of one LOD drop

around the marker was also given. A computer program takes each LOD peak and links it to the consensus through the markerhit in the database. The database is then queried for all HMM genes within the span of one LOD drop or a minimum of 15 Mb in each direction from the marker. The output is the name of the peak marker and its distance to the HMM gene.

The full sequences of the protease genes and splice variants are shown in Appendix I as SEQ ID NOs:1-47. The amino acids encoded by these nucleic acids are shown in Appendix I as SEQ ID NOs:48-94.

NUCLEIC ACIDS OF THE INVENTION

Protease Nucleic Acids, Portions and Variants

Accordingly, the invention pertains to isolated nucleic acid molecules comprising human protease genes. The term, "protease", as used herein, refers to an isolated nucleic acid molecule selected from the group shown in Appendix I, and consisting of SEQ ID NOs:1-47, and also to a portion or fragment of the isolated nucleic acid molecule (*e.g.*, cDNA or the gene) that encodes protease polypeptide (*e.g.*, a polypeptide selected from the group shown in Appendix I, and consisting of SEQ ID NOs:48-94). In a preferred embodiment, the isolated nucleic acid molecule comprises a nucleic acid molecule selected from the group consisting of SEQ ID NOs:1-47 or the complement of such a nucleic acid molecule.

The isolated nucleic acid molecules of the present invention can be RNA, for example, mRNA, or DNA, such as cDNA and genomic DNA. DNA molecules can be double-stranded or single-stranded; single stranded RNA or DNA can be either the coding, or sense, strand or the non-coding, or antisense, strand. The nucleic acid molecule can include all or a portion of the coding sequence of the gene and can further comprise additional non-coding sequences such as introns and non-coding 3' and 5' sequences (including regulatory sequences, for example). Additionally, the nucleic acid molecule can be fused to a marker sequence, for example, a sequence that encodes a polypeptide to assist in isolation or purification of the polypeptide. Such sequences include, but are not limited to, those which encode a

glutathione-S-transferase (GST) fusion protein and those which encode a hemagglutinin A (HA) polypeptide marker from influenza.

An "isolated" nucleic acid molecule, as used herein, is one that is separated from nucleic acids which normally flank the gene or nucleotide sequence (as in genomic sequences) and/or has been completely or partially purified from other transcribed sequences (*e.g.*, as in an RNA library). For example, an isolated nucleic acid of the invention may be substantially isolated with respect to the complex cellular milieu in which it naturally occurs, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. In some instances, the isolated material will form part of a composition (for example, a crude extract containing other substances), buffer system or reagent mix. In other circumstances, the material may be purified to essential homogeneity, for example as determined by PAGE or column chromatography such as HPLC. Preferably, an isolated nucleic acid molecule comprises at least about 50, 80 or 90% (on a molar basis) of all macromolecular species present. With regard to genomic DNA, the term "isolated" also can refer to nucleic acid molecules which are separated from the chromosome with which the genomic DNA is naturally associated. For example, the isolated nucleic acid molecule can contain less than about 5 kb, 4 kb, 3 kb, 2 kb, 1 kb, 0.5 kb or 0.1 kb of nucleotides which flank the nucleic acid molecule in the genomic DNA of the cell from which the nucleic acid molecule is derived.

The nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered isolated. Thus, recombinant DNA contained in a vector is included in the definition of "isolated" as used herein. Also, isolated nucleic acid molecules include recombinant DNA molecules in heterologous host cells, as well as partially or substantially purified DNA molecules in solution. "Isolated" nucleic acid molecules also encompass *in vivo* and *in vitro* RNA transcripts of the DNA molecules of the present invention. An isolated nucleic acid molecule or nucleotide sequence can include a nucleic acid molecule or nucleotide sequence which is synthesized chemically or by recombinant means. Therefore, recombinant DNA contained in a vector are included in the definition of "isolated"

as used herein. Also, isolated nucleotide sequences include recombinant DNA molecules in heterologous organisms, as well as partially or substantially purified DNA molecules in solution. *In vivo* and *in vitro* RNA transcripts of the DNA molecules of the present invention are also encompassed by "isolated" nucleotide sequences. Such isolated nucleotide sequences are useful in the manufacture of the encoded polypeptide, as probes for isolating homologous sequences (*e.g.*, from other mammalian species), for gene mapping (*e.g.*, by *in situ* hybridization with chromosomes), or for detecting expression of the gene in tissue (*e.g.*, human tissue), such as by Northern blot analysis.

The present invention also pertains to nucleic acid molecules which are not necessarily found in nature but which encode a protease polypeptide (*e.g.*, a polypeptide having an amino acid sequence comprising an amino acid sequence selected from the group consisting of SEQ ID NOs:48-94), or another splicing variant of a protease polypeptide or polymorphic variant thereof. Thus, for example, DNA molecules which comprise a sequence that is different from the naturally-occurring nucleotide sequence but which, due to the degeneracy of the genetic code, encode a protease polypeptide of the present invention are also the subject of this invention. The invention also encompasses nucleotide sequences encoding portions (fragments), or encoding variant polypeptides such as analogues or derivatives of a protease polypeptide. Such variants can be naturally-occurring, such as in the case of allelic variation or single nucleotide polymorphisms, or non-naturally-occurring, such as those induced by various mutagens and mutagenic processes. Intended variations include, but are not limited to, addition, deletion and substitution of one or more nucleotides which can result in conservative or non-conservative amino acid changes, including additions and deletions. Preferably the nucleotide (and/or resultant amino acid) changes are silent or conserved; that is, they do not alter the characteristics or activity of a protease polypeptide. In one preferred embodiment, the nucleotide sequences are fragments that comprise one or more polymorphic microsatellite markers. In another preferred embodiment, the nucleotide sequences are fragments that comprise one or more single nucleotide polymorphisms in a protease gene.

Other alterations of the nucleic acid molecules of the invention can include, for example, labeling, methylation, internucleotide modifications such as uncharged linkages (*e.g.*, methyl phosphonates, phosphotriesters, phosphoamidates, carbamates), charged linkages (*e.g.*, phosphorothioates, phosphorodithioates), pendent moieties (*e.g.*, polypeptides), intercalators (*e.g.*, acridine, psoralen), chelators, alkylators, and modified linkages (*e.g.*, alpha anomeric nucleic acids). Also included are synthetic molecules that mimic nucleic acid molecules in the ability to bind to a designated sequences via hydrogen bonding and other chemical interactions. Such molecules include, for example, those in which peptide linkages substitute for phosphate linkages in the backbone of the molecule.

The invention also pertains to nucleic acid molecules which hybridize under high stringency hybridization conditions, such as for selective hybridization, to a nucleotide sequence described herein (*e.g.*, nucleic acid molecules which specifically hybridize to a nucleotide sequence encoding polypeptides described herein, and, optionally, have an activity of the polypeptide). In one embodiment, the invention includes variants described herein which hybridize under high stringency hybridization conditions (*e.g.*, for selective hybridization) to a nucleotide sequence comprising a nucleotide sequence selected from the group consisting of SEQ ID NOs:1-47. In another embodiment, the invention includes variants described herein which hybridize under high stringency hybridization conditions (*e.g.*, for selective hybridization) to a nucleotide sequence encoding an amino acid sequence selected from the group consisting of SEQ ID NOs:48-94 or a polymorphic variant thereof. In a preferred embodiment, the variant which hybridizes under high stringency hybridizations has an activity of a protease.

Such nucleic acid molecules can be detected and/or isolated by specific hybridization (*e.g.*, under high stringency conditions). "Specific hybridization," as used herein, refers to the ability of a first nucleic acid to hybridize to a second nucleic acid in a manner such that the first nucleic acid does not hybridize to any nucleic acid other than to the second nucleic acid (*e.g.*, when the first nucleic acid has a higher similarity to the second nucleic acid than to any other nucleic acid in a sample wherein the hybridization is to be performed). "Stringency conditions" for

hybridization is a term of art which refers to the incubation and wash conditions, *e.g.*, conditions of temperature and buffer concentration, which permit hybridization of a particular nucleic acid to a second nucleic acid; the first nucleic acid may be perfectly (*i.e.*, 100%) complementary to the second, or the first and second may share some degree of complementarity which is less than perfect (*e.g.*, 70%, 75%, 85%, 95%). For example, certain high stringency conditions can be used which distinguish perfectly complementary nucleic acids from those of less complementarity. "High stringency conditions", "moderate stringency conditions" and "low stringency conditions" for nucleic acid hybridizations are explained on pages 2.10.1-2.10.16 and pages 6.3.1-6.3.6 in *Current Protocols in Molecular Biology* (Ausubel, F.M. *et al.*, "*Current Protocols in Molecular Biology*", John Wiley & Sons, (1998), the entire teachings of which are incorporated by reference herein). The exact conditions which determine the stringency of hybridization depend not only on ionic strength (*e.g.*, 0.2X SSC, 0.1X SSC), temperature (*e.g.*, room temperature, 42°C, 68°C) and the concentration of destabilizing agents such as formamide or denaturing agents such as SDS, but also on factors such as the length of the nucleic acid sequence, base composition, percent mismatch between hybridizing sequences and the frequency of occurrence of subsets of that sequence within other non-identical sequences. Thus, equivalent conditions can be determined by varying one or more of these parameters while maintaining a similar degree of identity or similarity between the two nucleic acid molecules. Typically, conditions are used such that sequences at least about 60%, at least about 70%, at least about 80%, at least about 90% or at least about 95% or more identical to each other remain hybridized to one another. By varying hybridization conditions from a level of stringency at which no hybridization occurs to a level at which hybridization is first observed, conditions which will allow a given sequence to hybridize (*e.g.*, selectively) with the most similar sequences in the sample can be determined.

Exemplary conditions are described in Krause, M.H. and S.A. Aaronson, *Methods in Enzymology* 200:546-556 (1991), and in, Ausubel, *et al.*, "*Current Protocols in Molecular Biology*", John Wiley & Sons, (1998), which describes the determination of washing conditions for moderate or low stringency conditions.

Washing is the step in which conditions are usually set so as to determine a minimum level of complementarity of the hybrids. Generally, starting from the lowest temperature at which only homologous hybridization occurs, each °C by which the final wash temperature is reduced (holding SSC concentration constant) allows an increase by 1% in the maximum extent of mismatching among the sequences that hybridize. Generally, doubling the concentration of SSC results in an increase in T_m of ~17°C. Using these guidelines, the washing temperature can be determined empirically for high, moderate or low stringency, depending on the level of mismatch sought.

For example, a low stringency wash can comprise washing in a solution containing 0.2X SSC/0.1% SDS for 10 minutes at room temperature; a moderate stringency wash can comprise washing in a prewarmed solution (42°C) solution containing 0.2X SSC/0.1% SDS for 15 minutes at 42°C; and a high stringency wash can comprise washing in prewarmed (68°C) solution containing 0.1X SSC/0.1%SDS for 15 minutes at 68°C. Furthermore, washes can be performed repeatedly or sequentially to obtain a desired result as known in the art. Equivalent conditions can be determined by varying one or more of the parameters given as an example, as known in the art, while maintaining a similar degree of identity or similarity between the target nucleic acid molecule and the primer or probe used.

The percent homology or identity of two nucleotide or amino acid sequences can be determined by aligning the sequences for optimal comparison purposes (*e.g.*, gaps can be introduced in the sequence of a first sequence for optimal alignment). The nucleotides or amino acids at corresponding positions are then compared, and the percent identity between the two sequences is a function of the number of identical positions shared by the sequences (*i.e.*, % identity = # of identical positions/total # of positions x 100). When a position in one sequence is occupied by the same nucleotide or amino acid residue as the corresponding position in the other sequence, then the molecules are homologous at that position. As used herein, nucleic acid or amino acid "homology" is equivalent to nucleic acid or amino acid "identity". In certain embodiments, the length of a sequence aligned for comparison purposes is at least 30%, for example, at least 40%, in certain embodiments at least

60%, and in other embodiments at least 70%, 80%, 90% or 95% of the length of the reference sequence. The actual comparison of the two sequences can be accomplished by well-known methods, for example, using a mathematical algorithm. A preferred, non-limiting example of such a mathematical algorithm is described in Karlin *et al.*, *Proc. Natl. Acad. Sci. USA* 90:5873-5877 (1993). Such an algorithm is incorporated into the NBLAST and XBLAST programs (version 2.0) as described in Altschul *et al.*, *Nucleic Acids Res.* 25:389-3402 (1997). When utilizing BLAST and Gapped BLAST programs, the default parameters of the respective programs (e.g., NBLAST) can be used. In one embodiment, parameters for sequence comparison can be set at score=100, wordlength=12, or can be varied (e.g., W=5 or W=20).

Another preferred, non-limiting example of a mathematical algorithm utilized for the comparison of sequences is the algorithm of Myers and Miller, *CABIOS* 4(1):11-17 (1988). Such an algorithm is incorporated into the ALIGN program (version 2.0) which is part of the GCG sequence alignment software package. When utilizing the ALIGN program for comparing amino acid sequences, a PAM120 weight residue table, a gap length penalty of 12 , and a gap penalty of 4 can be used. Additional algorithms for sequence analysis are known in the art and include ADVANCE and ADAM as described in Torellis and Robotti, *Comput. Appl. Biosci.* 10:3-5 (1994); and FASTA described in Pearson and Lipman, *Proc. Natl. Acad. Sci. USA* 85:2444-8 (1988).

In another embodiment, the percent identity between two amino acid sequences can be accomplished using the GAP program in the GCG software package (Accelrys, Cambridge, UK) using either a BLOSUM63 matrix or a PAM250 matrix, and a gap weight of 12, 10, 8, 6, or 4 and a length weight of 2, 3, or 4. In yet another embodiment, the percent identity between two nucleic acid sequences can be accomplished using the GAP program in the GCG software package, using a gap weight of 50 and a length weight of 3.

The present invention also provides isolated nucleic acid molecules that contain a fragment or portion that hybridizes under highly stringent conditions to a nucleotide sequence comprising a nucleotide sequence selected from the group

consisting of SEQ ID NOs:1-47, or the complement of such a sequence, and also provides isolated nucleic acid molecules that contain a fragment or portion that hybridizes under highly stringent conditions to a nucleotide sequence encoding an amino acid sequence selected SEQ ID NOs:48-94, or polymorphic variant thereof. The nucleic acid fragments of the invention are at least about 15, preferably at least about 18, 20, 23 or 25 nucleotides, and can be 30, 40, 50, 100, 200 or more nucleotides in length. Longer fragments, for example, 30 or more nucleotides in length, which encode antigenic polypeptides described herein are particularly useful, such as for the generation of antibodies as described below.

Probes and Primes

In a related aspect, the nucleic acid fragments of the invention are used as probes or primers in assays such as those described herein. "Probes" or "primers" are oligonucleotides that hybridize in a base-specific manner to a complementary strand of nucleic acid molecules. Such probes and primers include polypeptide nucleic acids, as described in Nielsen *et al.*, *Science* 254:1497-1500 (1991).

A probe or primer comprises a region of nucleotide sequence that hybridizes to at least about 15, typically about 20-25, and more typically about 40, 50 or 75, consecutive nucleotides of a nucleic acid molecule comprising a contiguous nucleotide sequence selected from the group consisting of SEQ ID NOs:1-47, or the complement of such a sequence, or a sequence encoding an amino acid sequence selected from SEQ ID NOs:48-94, or polymorphic variant thereof. In preferred embodiments, a probe or primer comprises 100 or fewer nucleotides, preferably from 6 to 50 nucleotides, preferably from 12 to 30 nucleotides. In other embodiments, the probe or primer is at least 70% identical to the contiguous nucleotide sequence or to the complement of the contiguous nucleotide sequence, preferably at least 80% identical, more preferably at least 90% identical, even more preferably at least 95% identical, or even capable of selectively hybridizing to the contiguous nucleotide sequence or to the complement of the contiguous nucleotide sequence. Often, the probe or primer further comprises a label, *e.g.*, radioisotope, fluorescent compound, enzyme, or enzyme co-factor.

The nucleic acid molecules of the invention such as those described above can be identified and isolated using standard molecular biology techniques and the sequence information provided herein. For example, nucleic acid molecules can be amplified and isolated by the polymerase chain reaction using synthetic oligonucleotide primers designed based on one or more of the sequences selected from the group consisting of SEQ ID NOs:1-47, or the complement of such a sequence, or designed based on nucleotides based on sequences encoding one or more of the amino acid sequences provided herein. See generally *PCR Technology: Principles and Applications for DNA Amplification* (ed. H.A. Erlich, Freeman Press, NY, NY, 1992); *PCR Protocols: A Guide to Methods and Applications* (Eds. Innis *et al.*, Academic Press, San Diego, CA, 1990); Mattila *et al.*, *Nucl. Acids Res.* 19:4967 (1991); Eckert *et al.*, *PCR Methods and Applications* 1:17 (1991); PCR (eds. McPherson *et al.*, IRL Press, Oxford); and U.S. Patent 4,683,202. The nucleic acid molecules can be amplified using cDNA, mRNA or genomic DNA as a template, cloned into an appropriate vector and characterized by DNA sequence analysis.

Other suitable amplification methods include the ligase chain reaction (LCR) (see Wu and Wallace, *Genomics* 4:560 (1989), Landegren *et al.*, *Science* 241:1077 (1988), transcription amplification (Kwoh *et al.*, *Proc. Natl. Acad. Sci. USA* 86:1173 (1989)), and self-sustained sequence replication (Guatelli *et al.*, *Proc. Nat. Acad. Sci. USA* 87:1874 (1990)) and nucleic acid based sequence amplification (NASBA). The latter two amplification methods involve isothermal reactions based on isothermal transcription, which produce both single stranded RNA (ssRNA) and double stranded DNA (dsDNA) as the amplification products in a ratio of about 30 or 100 to 1, respectively.

The amplified DNA can be labelled, for example radiolabelled, and used as a probe for screening a cDNA library derived from human cells, mRNA in zap express, ZIPLOX or other suitable vector. Corresponding clones can be isolated, DNA can obtained following *in vivo* excision, and the cloned insert can be sequenced in either or both orientations by art recognized methods to identify the correct reading frame encoding a polypeptide of the appropriate molecular weight. For example, the direct analysis of the nucleotide sequence of nucleic acid molecules

of the present invention can be accomplished using well-known methods that are commercially available. See, for example, Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual* (2nd Ed., CSHP, New York 1989); Zyskind *et al.*, *Recombinant DNA Laboratory Manual*, (Acad. Press, 1988)). Additionally, fluorescence methods are also available for analyzing nucleic acids (Chen *et al.*, *Genome Res.* 9, 492 (1999)) and polypeptides. Using these or similar methods, the polypeptide and the DNA encoding the polypeptide can be isolated, sequenced and further characterized.

Antisense nucleic acid molecules of the invention can be designed using the nucleotide sequences of one or more of SEQ ID NOs:1-47 and/or the complement of one or more of SEQ ID NOs:1-47, and/or a portion of one or more of SEQ ID NOs:1-47, or the complement of one or more of SEQ ID NOs:1-47 and/or a sequence encoding the amino acid sequences of one or more of SEQ ID NOs:48-94, or encoding a portion of one or more of SEQ ID NOs:48-94, and constructed using chemical synthesis and enzymatic ligation reactions using procedures known in the art. For example, an antisense nucleic acid molecule (*e.g.*, an antisense oligonucleotide) can be chemically synthesized using naturally occurring nucleotides or variously modified nucleotides designed to increase the biological stability of the molecules or to increase the physical stability of the duplex formed between the antisense and sense nucleic acids, *e.g.*, phosphorothioate derivatives and acridine substituted nucleotides can be used. Alternatively, the antisense nucleic acid molecule can be produced biologically using an expression vector into which a nucleic acid molecule has been subcloned in an antisense orientation (*i.e.*, RNA transcribed from the inserted nucleic acid molecule will be of an antisense orientation to a target nucleic acid of interest).

In general, the isolated nucleic acid sequences of the invention can be used as molecular weight markers on Southern gels, and as chromosome markers which are labeled to map related gene positions. The nucleic acid sequences can also be used to compare with endogenous DNA sequences in patients to identify one or more of the disorders described above, and as probes, such as to hybridize and discover related DNA sequences or to subtract out known sequences from a sample. The nucleic acid sequences can further be used to derive primers for genetic

fingerprinting, to raise anti-polypeptide antibodies using DNA immunization techniques, and as an antigen to raise anti-DNA antibodies or elicit immune responses. Portions or fragments of the nucleotide sequences identified herein (and the corresponding complete gene sequences) can be used in numerous ways as polynucleotide reagents. For example, these sequences can be used to: (i) map their respective genes on a chromosome; and, thus, locate gene regions associated with genetic disease; (ii) identify an individual from a minute biological sample (tissue typing); and (iii) aid in forensic identification of a biological sample. Additionally, the nucleotide sequences of the invention can be used to identify and express recombinant polypeptides for analysis, characterization or therapeutic use, or as markers for tissues in which the corresponding polypeptide is expressed, either constitutively, during tissue differentiation, or in diseased states. The nucleic acid sequences can additionally be used as reagents in the screening and/or diagnostic assays described herein, and can also be included as components of kits (*e.g.*, reagent kits) for use in the screening and/or diagnostic assays described herein.

Vectors

Another aspect of the invention pertains to nucleic acid constructs containing a nucleic acid molecule selected from the group consisting of SEQ ID NOs:1-47 and the complements thereof (or a portion thereof). Yet another aspect of the invention pertains to nucleic acid constructs containing a nucleic acid molecule encoding an amino acid sequence of SEQ ID NOs:48-94 or polymorphic variant thereof. The constructs comprise a vector (*e.g.*, an expression vector) into which a sequence of the invention has been inserted in a sense or antisense orientation. As used herein, the term "vector" refers to a nucleic acid molecule capable of transporting another nucleic acid to which it has been linked. One type of vector is a "plasmid", which refers to a circular double stranded DNA loop into which additional DNA segments can be ligated. Another type of vector is a viral vector, wherein additional DNA segments can be ligated into the viral genome. Certain vectors are capable of autonomous replication in a host cell into which they are introduced (*e.g.*, bacterial vectors having a bacterial origin of replication and episomal mammalian vectors).

Other vectors (*e.g.*, non-episomal mammalian vectors) are integrated into the genome of a host cell upon introduction into the host cell, and thereby are replicated along with the host genome. Moreover, certain vectors, expression vectors, are capable of directing the expression of genes to which they are operably linked. In general, expression vectors of utility in recombinant DNA techniques are often in the form of plasmids. However, the invention is intended to include such other forms of expression vectors, such as viral vectors (*e.g.*, replication defective retroviruses, adenoviruses and adeno-associated viruses) that serve equivalent functions.

Preferred recombinant expression vectors of the invention comprise a nucleic acid molecule of the invention in a form suitable for expression of the nucleic acid molecule in a host cell. This means that the recombinant expression vectors include one or more regulatory sequences, selected on the basis of the host cells to be used for expression, which is operably linked to the nucleic acid sequence to be expressed. Within a recombinant expression vector, "operably linked" or "operatively linked" is intended to mean that the nucleotide sequence of interest is linked to the regulatory sequence(s) in a manner which allows for expression of the nucleotide sequence (*e.g.*, in an *in vitro* transcription/translation system or in a host cell when the vector is introduced into the host cell). The term "regulatory sequence" is intended to include promoters, enhancers and other expression control elements (*e.g.*, polyadenylation signals). Such regulatory sequences are described, for example, in Goeddel, "Gene Expression Technology", *Methods in Enzymology* 185, Academic Press, San Diego, CA (1990). Regulatory sequences include those which direct constitutive expression of a nucleotide sequence in many types of host cell and those which direct expression of the nucleotide sequence only in certain host cells (*e.g.*, tissue-specific regulatory sequences). It will be appreciated by those skilled in the art that the design of the expression vector can depend on such factors as the choice of the host cell to be transformed and the level of expression of polypeptide desired. The expression vectors of the invention can be introduced into host cells to thereby produce polypeptides, including fusion polypeptides, encoded by nucleic acid molecules as described herein.

The recombinant expression vectors of the invention can be designed for expression of a polypeptide of the invention in prokaryotic or eukaryotic cells, *e.g.*, bacterial cells such as *E. coli*, insect cells (using baculovirus expression vectors), yeast cells or mammalian cells. Suitable host cells are discussed further in Goeddel, *supra*. Alternatively, the recombinant expression vector can be transcribed and translated *in vitro*, for example using T7 promoter regulatory sequences and T7 polymerase.

Another aspect of the invention pertains to host cells into which a recombinant expression vector of the invention has been introduced. The terms "host cell" and "recombinant host cell" are used interchangeably herein. It is understood that such terms refer not only to the particular subject cell but also to the progeny or potential progeny of such a cell. Because certain modifications may occur in succeeding generations due to either mutation or environmental influences, such progeny may not, in fact, be identical to the parent cell, but are still included within the scope of the term as used herein.

A host cell can be any prokaryotic or eukaryotic cell. For example, a nucleic acid molecule of the invention can be expressed in bacterial cells (*e.g.*, *E. coli*), insect cells, yeast or mammalian cells (such as Chinese hamster ovary cells (CHO) or COS cells). Other suitable host cells are known to those skilled in the art.

Vector DNA can be introduced into prokaryotic or eukaryotic cells via conventional transformation or transfection techniques. As used herein, the terms "transformation" and "transfection" are intended to refer to a variety of art-recognized techniques for introducing a foreign nucleic acid molecule (*e.g.*, DNA) into a host cell, including calcium phosphate or calcium chloride co-precipitation, DEAE-dextran-mediated transfection, lipofection, or electroporation. Suitable methods for transforming or transfecting host cells can be found in Sambrook, *et al.* (*supra*), and other laboratory manuals.

For stable transfection of mammalian cells, it is known that, depending upon the expression vector and transfection technique used, only a small fraction of cells may integrate the foreign DNA into their genome. In order to identify and select these integrants, a gene that encodes a selectable marker (*e.g.*, for resistance to

antibiotics) is generally introduced into the host cells along with the gene of interest. Preferred selectable markers include those that confer resistance to drugs, such as G418, hygromycin and methotrexate. Nucleic acid molecules encoding a selectable marker can be introduced into a host cell on the same vector as the nucleic acid molecule of the invention or can be introduced on a separate vector. Cells stably transfected with the introduced nucleic acid molecule can be identified by drug selection (*e.g.*, cells that have incorporated the selectable marker gene will survive, while the other cells die).

A host cell of the invention, such as a prokaryotic or eukaryotic host cell in culture, can be used to produce (*i.e.*, express) a polypeptide of the invention. Accordingly, the invention further provides methods for producing a polypeptide using the host cells of the invention. In one embodiment, the method comprises culturing the host cell of invention (into which a recombinant expression vector encoding a polypeptide of the invention has been introduced) in a suitable medium such that the polypeptide is produced. In another embodiment, the method further comprises isolating the polypeptide from the medium or the host cell.

The host cells of the invention can also be used to produce nonhuman transgenic animals. For example, in one embodiment, a host cell of the invention is a fertilized oocyte or an embryonic stem cell into which a nucleic acid molecule of the invention has been introduced (*e.g.*, an exogenous protease gene, or an exogenous nucleic acid encoding a protease polypeptide). Such host cells can then be used to create non-human transgenic animals in which exogenous nucleotide sequences have been introduced into the genome or homologous recombinant animals in which endogenous nucleotide sequences have been altered. Such animals are useful for studying the function and/or activity of the nucleotide sequence and polypeptide encoded by the sequence and for identifying and/or evaluating modulators of their activity. As used herein, a "transgenic animal" is a non-human animal, preferably a mammal, more preferably a rodent such as a rat or mouse, in which one or more of the cells of the animal includes a transgene. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens and amphibians. A transgene is exogenous DNA which is integrated into

the genome of a cell from which a transgenic animal develops and which remains in the genome of the mature animal, thereby directing the expression of an encoded gene product in one or more cell types or tissues of the transgenic animal. As used herein, an "homologous recombinant animal" is a non-human animal, preferably a mammal, more preferably a mouse, in which an endogenous gene has been altered by homologous recombination between the endogenous gene and an exogenous DNA molecule introduced into a cell of the animal, *e.g.*, an embryonic cell of the animal, prior to development of the animal.

Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, U.S. Pat. No. 4,873,191 and in Hogan, *Manipulating the Mouse Embryo* (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Methods for constructing homologous recombination vectors and homologous recombinant animals are described further in Bradley, *Current Opinion in BioTechnology* 2:823-829 (1991) and in PCT Publication Nos. WO 90/11354, WO 91/01140, WO 92/0968, and WO 93/04169. Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut *et al.*, *Nature* 385:810-813 (1997) and PCT Publication Nos. WO 97/07668 and WO 97/07669.

POLYPEPTIDES OF THE INVENTION

The present invention also pertains to isolated polypeptides encoded by proteases ("protease polypeptides") and fragments and variants thereof, as well as polypeptides encoded by nucleotide sequences described herein (*e.g.*, other splicing variants). The term "polypeptide" refers to a polymer of amino acids, and not to a specific length; thus, peptides, oligopeptides and proteins are included within the definition of a polypeptide. As used herein, a polypeptide is said to be "isolated" or "purified" when it is substantially free of cellular material when it is isolated from recombinant and non-recombinant cells, or free of chemical precursors or other chemicals when it is chemically synthesized. A polypeptide, however, can be joined

to another polypeptide with which it is not normally associated in a cell (*e.g.*, in a "fusion protein") and still be "isolated" or "purified."

The polypeptides of the invention can be purified to homogeneity. It is understood, however, that preparations in which the polypeptide is not purified to homogeneity are useful. The critical feature is that the preparation allows for the desired function of the polypeptide, even in the presence of considerable amounts of other components. Thus, the invention encompasses various degrees of purity. In one embodiment, the language "substantially free of cellular material" includes preparations of the polypeptide having less than about 30% (by dry weight) other proteins (*i.e.*, contaminating protein), less than about 20% other proteins, less than about 10% other proteins, or less than about 5% other proteins.

When a polypeptide is recombinantly produced, it can also be substantially free of culture medium, *i.e.*, culture medium represents less than about 20%, less than about 10%, or less than about 5% of the volume of the polypeptide preparation. The language "substantially free of chemical precursors or other chemicals" includes preparations of the polypeptide in which it is separated from chemical precursors or other chemicals that are involved in its synthesis. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of the polypeptide having less than about 30% (by dry weight) chemical precursors or other chemicals, less than about 20% chemical precursors or other chemicals, less than about 10% chemical precursors or other chemicals, or less than about 5% chemical precursors or other chemicals.

In one embodiment, a polypeptide of the invention comprises an amino acid sequence encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOs:1-47, or the complement of such a nucleic acid, or portions thereof, *e.g.*, SEQ ID NO:48-94, or a portion or polymorphic variant thereof. However, the polypeptides of the invention also encompass fragment and sequence variants. Variants include a substantially homologous polypeptide encoded by the same genetic locus in an organism, *i.e.*, an allelic variant, as well as other splicing variants. Variants also encompass polypeptides derived from other genetic loci in an organism, but having substantial

homology to a polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of SEQ ID NOs:1-47, or a complement of such a sequence, or portions thereof, or having substantial homology to a polypeptide encoded by a nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of nucleotide sequences encoding SEQ ID NOs:48-94, or polymorphic variants thereof. Variants also include polypeptides substantially homologous or identical to these polypeptides but derived from another organism, *i.e.*, an ortholog. Variants also include polypeptides that are substantially homologous or identical to these polypeptides that are produced by chemical synthesis. Variants also include polypeptides that are substantially homologous or identical to these polypeptides that are produced by recombinant methods.

As used herein, two polypeptides (or a region of the polypeptides) are substantially homologous or identical when the amino acid sequences are at least about 45-55%, in certain embodiments at least about 70-75%, in other embodiments at least about 80-85%, and in other embodiments greater than about 90% or more homologous or identical. A substantially homologous amino acid sequence, according to the present invention, will be encoded by a nucleic acid molecule hybridizing to one or more of SEQ ID NOs:1-47, or portion thereof, under stringent conditions as more particularly described above, or will be encoded by a nucleic acid molecule hybridizing to a nucleic acid sequence encoding one of SEQ ID NOs:48-94, a portion thereof or polymorphic variant thereof, under stringent conditions as more particularly described thereof.

The invention also encompasses polypeptides having a lower degree of identity but having sufficient similarity so as to perform one or more of the same functions performed by a polypeptide encoded by a nucleic acid molecule of the invention. Similarity is determined by conserved amino acid substitution. Such substitutions are those that substitute a given amino acid in a polypeptide by another amino acid of like characteristics. Conservative substitutions are likely to be phenotypically silent. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu and Ile; interchange of the hydroxyl residues Ser and Thr, exchange of the acidic residues

Asp and Glu, substitution between the amide residues Asn and Gln, exchange of the basic residues Lys and Arg and replacements among the aromatic residues Phe and Tyr. Guidance concerning which amino acid changes are likely to be phenotypically silent are found in Bowie *et al.*, *Science* 247:1306-1310 (1990).

A variant polypeptide can differ in amino acid sequence by one or more substitutions, deletions, insertions, inversions, fusions, and truncations or a combination of any of these. Further, variant polypeptides can be fully functional or can lack function in one or more activities. Fully functional variants typically contain only conservative variation or variation in non-critical residues or in non-critical regions. Functional variants can also contain substitution of similar amino acids that result in no change or an insignificant change in function. Alternatively, such substitutions may positively or negatively affect function to some degree. Non-functional variants typically contain one or more non-conservative amino acid substitutions, deletions, insertions, inversions, or truncation or a substitution, insertion, inversion, or deletion in a critical residue or critical region.

Amino acids that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham *et al.*, *Science* 244:1081-1085 (1989)). The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity *in vitro*, or *in vitro* proliferative activity. Sites that are critical for polypeptide activity can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling (Smith *et al.*, *J. Mol. Biol. 224*:899-904 (1992); de Vos *et al.*, *Science 255*:306-312 (1992)).

The invention also includes polypeptide fragments of the polypeptides of the invention. Fragments can be derived from a polypeptide encoded by a nucleic acid molecule comprising one of SEQ ID NOs:1-47, or a complement of such a nucleic acid (*e.g.*, SEQ ID NOs:48-94, or other variants). However, the invention also encompasses fragments of the variants of the polypeptides described herein. As used herein, a fragment comprises at least 6 contiguous amino acids. Useful fragments include those that retain one or more of the biological activities of the

polypeptide as well as fragments that can be used as an immunogen to generate polypeptide-specific antibodies.

Biologically active fragments (peptides which are, for example, 6, 9, 12, 15, 16, 20, 30, 35, 36, 37, 38, 39, 40, 50, 100 or more amino acids in length) can comprise a domain, segment, or motif that has been identified by analysis of the polypeptide sequence using well-known methods, *e.g.*, signal peptides, extracellular domains, one or more transmembrane segments or loops, ligand binding regions, zinc finger domains, DNA binding domains, acylation sites, glycosylation sites, or phosphorylation sites.

Fragments can be discrete (not fused to other amino acids or polypeptides) or can be within a larger polypeptide. Further, several fragments can be comprised within a single larger polypeptide. In one embodiment a fragment designed for expression in a host can have heterologous pre- and pro-polypeptide regions fused to the amino terminus of the polypeptide fragment and an additional region fused to the carboxyl terminus of the fragment.

The invention thus provides chimeric or fusion polypeptides. These comprise a polypeptide of the invention operatively linked to a heterologous protein or polypeptide having an amino acid sequence not substantially homologous to the polypeptide. "Operatively linked" indicates that the polypeptide and the heterologous protein are fused in-frame. The heterologous protein can be fused to the N-terminus or C-terminus of the polypeptide. In one embodiment the fusion polypeptide does not affect function of the polypeptide *per se*. For example, the fusion polypeptide can be a GST-fusion polypeptide in which the polypeptide sequences are fused to the C-terminus of the GST sequences. Other types of fusion polypeptides include, but are not limited to, enzymatic fusion polypeptides, for example β-galactosidase fusions, yeast two-hybrid GAL fusions, poly-His fusions and Ig fusions. Such fusion polypeptides, particularly poly-His fusions, can facilitate the purification of recombinant polypeptide. In certain host cells (*e.g.*, mammalian host cells), expression and/or secretion of a polypeptide can be increased by using a heterologous signal sequence. Therefore, in another

embodiment, the fusion polypeptide contains a heterologous signal sequence at its N-terminus.

EP A 0 464 533 discloses fusion proteins comprising various portions of immunoglobulin constant regions. The Fc is useful in therapy and diagnosis and thus results, for example, in improved pharmacokinetic properties (EP A 0 232 262). In drug discovery, for example, human proteins have been fused with Fc portions for the purpose of high-throughput screening assays to identify antagonists. Bennett *et al.*, *Journal of Molecular Recognition, 8*:52-58 (1995) and Johanson *et al.*, *The Journal of Biological Chemistry, 270*,16:9459-9471 (1995). Thus, this invention also encompasses soluble fusion polypeptides containing a polypeptide of the invention and various portions of the constant regions of heavy or light chains of immunoglobulins of various subclass (IgG, IgM, IgA, IgE).

A chimeric or fusion polypeptide can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different polypeptide sequences are ligated together in-frame in accordance with conventional techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of nucleic acid fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive nucleic acid fragments which can subsequently be annealed and re-amplified to generate a chimeric nucleic acid sequence (see Ausubel *et al.*, *Current Protocols in Molecular Biology*, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (*e.g.*, a GST protein). A nucleic acid molecule encoding a polypeptide of the invention can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the polypeptide.

The isolated polypeptide can be purified from cells that naturally express it, purified from cells that have been altered to express it (recombinant), or synthesized using known protein synthesis methods. In one embodiment, the polypeptide is produced by recombinant DNA techniques. For example, a nucleic acid molecule encoding the polypeptide is cloned into an expression vector, the expression vector introduced into a host cell and the polypeptide expressed in the host cell. The

polypeptide can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques.

In general, polypeptides of the present invention can be used as a molecular weight marker on SDS-PAGE gels or on molecular sieve gel filtration columns using art-recognized methods. The polypeptides of the present invention can be used to raise antibodies or to elicit an immune response. The polypeptides can also be used as a reagent, *e.g.*, a labeled reagent, in assays to quantitatively determine levels of the polypeptide or a molecule to which it binds (*e.g.*, a ligand) in biological fluids. The polypeptides can also be used as markers for cells or tissues in which the corresponding polypeptide is preferentially expressed, either constitutively, during tissue differentiation, or in a diseased state. The polypeptides can be used to isolate a corresponding binding agent, *e.g.*, ligand, such as, for example, in an interaction trap assay, and to screen for peptide or small molecule antagonists or agonists of the binding interaction.

ANTIBODIES OF THE INVENTION

Polyclonal and/or monoclonal antibodies that specifically bind one form of the gene product but not to the other form of the gene product are also provided. Antibodies are also provided that bind a portion of either the variant or the reference gene product that contains the polymorphic site or sites. The invention provides antibodies to the polypeptides and polypeptide fragments of the invention, *e.g.*, having an amino acid sequence of one of SEQ ID NOs:48-94 or a portion thereof, or having an amino acid sequence encoded by a nucleic acid molecule comprising all or a portion of SEQ ID NOs:1-47, or a complement or another variant or portion thereof. The term "antibody" as used herein refers to immunoglobulin molecules and immunologically active portions of immunoglobulin molecules, *i.e.*, molecules that contain an antigen binding site that specifically binds an antigen. A molecule that specifically binds to a polypeptide of the invention is a molecule that binds to that polypeptide or a fragment thereof, but does not substantially bind other molecules in a sample, *e.g.*, a biological sample, which naturally contains the polypeptide. Examples of immunologically active portions of immunoglobulin

molecules include F(ab) and F(ab')$_2$ fragments which can be generated by treating the antibody with an enzyme such as pepsin. The invention provides polyclonal and monoclonal antibodies that bind to a polypeptide of the invention. The term "monoclonal antibody" or "monoclonal antibody composition", as used herein, refers to a population of antibody molecules that contain only one species of an antigen binding site capable of immunoreacting with a particular epitope of a polypeptide of the invention. A monoclonal antibody composition thus typically displays a single binding affinity for a particular polypeptide of the invention with which it immunoreacts.

Polyclonal antibodies can be prepared as described above by immunizing a suitable subject with a desired immunogen, *e.g.*, polypeptide of the invention or fragment thereof. The antibody titer in the immunized subject can be monitored over time by standard techniques, such as with an enzyme linked immunosorbent assay (ELISA) using immobilized polypeptide. If desired, the antibody molecules directed against the polypeptide can be isolated from the mammal (*e.g.*, from the blood) and further purified by well-known techniques, such as protein A chromatography to obtain the IgG fraction. At an appropriate time after immunization, *e.g.*, when the antibody titers are highest, antibody-producing cells can be obtained from the subject and used to prepare monoclonal antibodies by standard techniques, such as the hybridoma technique originally described by Kohler and Milstein, *Nature* 256:495-497 (1975), the human B cell hybridoma technique (Kozbor *et al.*, *Immunol. Today* 4:72 (1983)), the EBV-hybridoma technique (Cole *et al.*, *Monoclonal Antibodies and Cancer Therapy*, Alan R. Liss, 1985, Inc., pp. 77-96) or trioma techniques. The technology for producing hybridomas is well known (see generally *Current Protocols in Immunology* (1994) Coligan *et al.* (eds.) John Wiley & Sons, Inc., New York, NY). Briefly, an immortal cell line (typically a myeloma) is fused to lymphocytes (typically splenocytes) from a mammal immunized with an immunogen as described above, and the culture supernatants of the resulting hybridoma cells are screened to identify a hybridoma producing a monoclonal antibody that binds a polypeptide of the invention.

Any of the many well known protocols used for fusing lymphocytes and immortalized cell lines can be applied for the purpose of generating a monoclonal antibody to a polypeptide of the invention (see, *e.g.*, *Current Protocols in Immunology, supra*; Galfre *et al.*, *Nature* 266:55052 (1977); R.H. Kenneth, in *Monoclonal Antibodies: A New Dimension In Biological Analyses*, Plenum Publishing Corp., New York, New York (1980); and Lerner, *Yale J. Biol. Med.* 54:387-402 (1981). Moreover, the ordinarily skilled worker will appreciate that there are many variations of such methods that also would be useful.

Alternative to preparing monoclonal antibody-secreting hybridomas, a monoclonal antibody to a polypeptide of the invention can be identified and isolated by screening a recombinant combinatorial immunoglobulin library (*e.g.*, an antibody phage display library) with the polypeptide to thereby isolate immunoglobulin library members that bind the polypeptide. Kits for generating and screening phage display libraries are commercially available (*e.g.*, the Pharmacia *Recombinant Phage Antibody System*, Catalog No. 27-9400-01; and the Stratagene *Surf*ZAP™ Phage Display Kit, Catalog No. 240612). Additionally, examples of methods and reagents particularly amenable for use in generating and screening antibody display library can be found in, for example, U.S. Patent No. 5,223,409; PCT Publication No. WO 92/18619; PCT Publication No. WO 91/17271; PCT Publication No. WO 92/20791; PCT Publication No. WO 92/15679; PCT Publication No. WO 93/01288; PCT Publication No. WO 92/01047; PCT Publication No. WO 92/09690; PCT Publication No. WO 90/02809; Fuchs *et al.*, *Bio/Technology* 9:1370-1372 (1991); Hay *et al.*, *Hum. Antibod. Hybridomas* 3:81-85 (1992); Huse *et al.*, *Science* 246:1275-1281 (1989); Griffiths *et al.*, *EMBO J.* 12:725-734 (1993).

Additionally, recombinant antibodies, such as chimeric and humanized monoclonal antibodies, comprising both human and non-human portions, which can be made using standard recombinant DNA techniques, are within the scope of the invention. Such chimeric and humanized monoclonal antibodies can be produced by recombinant DNA techniques known in the art.

In general, antibodies of the invention (*e.g.*, a monoclonal antibody) can be used to isolate a polypeptide of the invention by standard techniques, such as affinity

chromatography or immunoprecipitation. A polypeptide-specific antibody can facilitate the purification of natural polypeptide from cells and of recombinantly produced polypeptide expressed in host cells. Moreover, an antibody specific for a polypeptide of the invention can be used to detect the polypeptide (*e.g.*, in a cellular lysate, cell supernatant, or tissue sample) in order to evaluate the abundance and pattern of expression of the polypeptide. Antibodies can be used diagnostically to monitor protein levels in tissue as part of a clinical testing procedure, *e.g.*, to, for example, determine the efficacy of a given treatment regimen. Detection can be facilitated by coupling the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

DIAGNOSTIC AND SCREENING ASSAYS OF THE INVENTION

The present invention also pertains to a method of diagnosing or aiding in the diagnosis of a disease or condition associated with a protease gene or gene product in an individual. Diagnostic assays can be designed for assessing protease gene expression, or for assessing activity of protease polypeptides of the invention. In one embodiment, the assays are used in the context of a biological sample (*e.g.*, blood, serum, cells, tissue) to thereby determine whether an individual is afflicted with a disease or condition associated with a protease, or a defect in a protease. The invention also provides for prognostic (or predictive) assays for determining whether an individual is susceptible to a disease or condition associated with a protease. For example, mutations in the gene can be assayed in a biological sample. Such assays

can be used for prognostic or predictive purpose to thereby prophylactically treat an individual prior to the onset of symptoms associated with a susceptibility to a disease or condition associated with a protease. Another aspect of the invention pertains to assays for monitoring the influence of agents (*e.g.*, drugs, compounds or other agents) on the gene expression or activity of polypeptides of the invention, as well as to assays for identifying agents which bind to a polypeptides. These and other assays and agents are described in further detail in the following sections.

DIAGNOSTIC ASSAYS

The nucleic acids, probes, primers, polypeptides and antibodies described herein can be used in methods of diagnosis of a susceptibility to a disease or condition associated with a protease, as well as in kits useful for diagnosis of a susceptibility to a disease or condition associated with a protease.

In one embodiment of the invention, diagnosis of a susceptibility to a disease or condition associated with a protease is made by detecting a polymorphism in a protease as described herein. The polymorphism can be a mutation in a protease, such as the insertion or deletion of a single nucleotide, or of more than one nucleotide, resulting in a frame shift mutation; the change of at least one nucleotide, resulting in a change in the encoded amino acid; the change of at least one nucleotide, resulting in the generation of a premature stop codon; the deletion of several nucleotides, resulting in a deletion of one or more amino acids encoded by the nucleotides; the insertion of one or several nucleotides, such as by unequal recombination or gene conversion, resulting in an interruption of the coding sequence of the gene; duplication of all or a part of the gene; transposition of all or a part of the gene; or rearrangement of all or a part of the gene. More than one such mutation may be present in a single gene. Such sequence changes cause a mutation in the polypeptide encoded by a protease gene. For example, if the mutation is a frame shift mutation, the frame shift can result in a change in the encoded amino acids, and/or can result in the generation of a premature stop codon, causing generation of a truncated polypeptide. Alternatively, a polymorphism associated with a susceptibility to a disease or condition associated with a protease can be a

synonymous alteration in one or more nucleotides (*i.e.*, an alteration that does not result in a change in the polypeptide encoded by a protease gene). Such a polymorphism may alter splicing sites, affect the stability or transport of mRNA, or otherwise affect the transcription or translation of the gene. A protease gene that has any of the alterations described above is referred to herein as a "altered gene."

In a first method of diagnosing a disease or a susceptibility to a disease or condition associated with a protease, hybridization methods, such as Southern analysis, Northern analysis, or *in situ* hybridizations, can be used (see *Current Protocols in Molecular Biology*, Ausubel, F. *et al.*, eds., John Wiley & Sons, including all supplements through 2001). For example, a biological sample from a test subject (a "test sample") of genomic DNA, RNA, or cDNA, is obtained from an individual suspected of having, being susceptible to or predisposed for, or carrying a defect for, a susceptibility to a disease or condition associated with a protease (the "test individual"). The individual can be an adult, child, or fetus. The test sample can be from any source which contains genomic DNA, such as a blood sample, sample of amniotic fluid, sample of cerebrospinal fluid, or tissue sample from skin, muscle, buccal or conjunctival mucosa, placenta, gastrointestinal tract or other organs. A test sample of DNA from fetal cells or tissue can be obtained by appropriate methods, such as by amniocentesis or chorionic villus sampling. The DNA, RNA, or cDNA sample is then examined to determine whether a polymorphism in a protease is present, and/or to determine which splicing variant(s) encoded by the protease is present. The presence of the polymorphism or splicing variant(s) can be indicated by hybridization of the gene in the genomic DNA, RNA, or cDNA to a nucleic acid probe. A "nucleic acid probe", as used herein, can be a DNA probe or an RNA probe; the nucleic acid probe can contain at least one polymorphism in a protease or contains a nucleic acid encoding a particular splicing variant of a protease. The probe can be any of the nucleic acid molecules described above (*e.g.*, the gene, a fragment, a vector comprising the gene, a probe or primer, etc.).

To diagnose a disease or a susceptibility to a disease or condition associated with a protease, a hybridization sample is formed by contacting the test sample

containing a protease, with at least one nucleic acid probe. A preferred probe for detecting mRNA or genomic DNA is a labeled nucleic acid probe capable of hybridizing to mRNA or genomic DNA sequences described herein. The nucleic acid probe can be, for example, a full-length nucleic acid molecule, or a portion thereof, such as an oligonucleotide of at least 15, 30, 50, 100, 250 or 500 nucleotides in length and sufficient to specifically hybridize under stringent conditions to appropriate mRNA or genomic DNA. For example, the nucleic acid probe can be all or a portion of one of SEQ ID NOs:1-47, or the complement thereof, or a portion thereof; or can be a nucleic acid encoding a portion of one of SEQ ID NOs:48-94. Other suitable probes for use in the diagnostic assays of the invention are described above (see *e.g.*, probes and primers discussed under the heading, "Nucleic Acids of the Invention").

The hybridization sample is maintained under conditions which are sufficient to allow specific hybridization of the nucleic acid probe to a protease. "Specific hybridization", as used herein, indicates exact hybridization (*e.g.*, with no mismatches). Specific hybridization can be performed under high stringency conditions or moderate stringency conditions, for example, as described above. In a particularly preferred embodiment, the hybridization conditions for specific hybridization are high stringency.

Specific hybridization, if present, is then detected using standard methods. If specific hybridization occurs between the nucleic acid probe and the protease in the test sample, then the protease has the polymorphism, or is the splicing variant, that is present in the nucleic acid probe. More than one nucleic acid probe can also be used concurrently in this method. Specific hybridization of any one of the nucleic acid probes is indicative of a polymorphism in the protease, or of the presence of a particular splicing variant encoding the protease and is therefore diagnostic for a disease or a susceptibility to a disease or condition associated with a protease.

In Northern analysis (see *Current Protocols in Molecular Biology*, Ausubel, F. *et al.*, eds., John Wiley & Sons, *supra*) the hybridization methods described above are used to identify the presence of a polymorphism or a particular splicing variant, associated with a disease or a susceptibility to a disease or condition associated with

a protease. For Northern analysis, a test sample of RNA is obtained from the individual by appropriate means. Specific hybridization of a nucleic acid probe, as described above, to RNA from the individual is indicative of a polymorphism in a protease, or of the presence of a particular splicing variant encoded by a protease, and is therefore diagnostic for a disease or a susceptibility to a disease or condition associated with a protease.

For representative examples of use of nucleic acid probes, see, for example, U.S. Patents No. 5,288,611 and 4,851,330.

Alternatively, a peptide nucleic acid (PNA) probe can be used instead of a nucleic acid probe in the hybridization methods described above. PNA is a DNA mimic having a peptide-like, inorganic backbone, such as N-(2-aminoethyl)glycine units, with an organic base (A, G, C, T or U) attached to the glycine nitrogen via a methylene carbonyl linker (see, for example, Nielsen, P.E. *et al.*, *Bioconjugate Chemistry* 5, American Chemical Society, p. 1 (1994). The PNA probe can be designed to specifically hybridize to a gene having a polymorphism associated with a disease or a susceptibility to a disease or condition associated with a protease. Hybridization of the PNA probe to a protease is diagnostic for a disease or a susceptibility to a disease or condition associated with a protease.

In another method of the invention, mutation analysis by restriction digestion can be used to detect a mutant gene, or genes containing a polymorphism(s), if the mutation or polymorphism in the gene results in the creation or elimination of a restriction site. A test sample containing genomic DNA is obtained from the individual. Polymerase chain reaction (PCR) can be used to amplify a protease (and, if necessary, the flanking sequences) in the test sample of genomic DNA from the test individual. RFLP analysis is conducted as described (see *Current Protocols in Molecular Biology*, *supra*). The digestion pattern of the relevant DNA fragment indicates the presence or absence of the mutation or polymorphism in the protease, and therefore indicates the presence or absence of this disease or a susceptibility to a disease or condition associated with a protease.

Sequence analysis can also be used to detect specific polymorphisms in a protease. A test sample of DNA or RNA is obtained from the test individual. PCR

or other appropriate methods can be used to amplify the gene, and/or its flanking sequences, if desired. The sequence of a protease, or a fragment of the gene, or cDNA, or fragment of the cDNA, or mRNA, or fragment of the mRNA, is determined, using standard methods. The sequence of the gene, gene fragment, cDNA, cDNA fragment, mRNA, or mRNA fragment is compared with the known nucleic acid sequence of the gene, cDNA (*e.g.*, one or more of SEQ ID NOs:1-47, or a complement thereof, or a nucleic acid sequence encoding one of SEQ ID NOs:48-94 or a fragment thereof) or mRNA, as appropriate. The presence of a polymorphism in the protease indicates that the individual has a disease or a susceptibility to a disease or condition associated with a protease.

Allele-specific oligonucleotides can also be used to detect the presence of a polymorphism in a protease, through the use of dot-blot hybridization of amplified oligonucleotides with allele-specific oligonucleotide (ASO) probes (see, for example, Saiki, R. *et al.*, *Nature* 324:163-166 (1986)). An "allele-specific oligonucleotide" (also referred to herein as an "allele-specific oligonucleotide probe") is an oligonucleotide of approximately 10-50 base pairs, preferably approximately 15-30 base pairs, that specifically hybridizes to a protease, and that contains a polymorphism associated with a disease or a susceptibility to a disease or condition associated with a protease. An allele-specific oligonucleotide probe that is specific for particular polymorphisms in a protease can be prepared, using standard methods (see *Current Protocols in Molecular Biology, supra*). To identify polymorphisms in the gene that are associated with a disease or a susceptibility to a disease or condition associated with a protease, a test sample of DNA is obtained from the individual. PCR can be used to amplify all or a fragment of a protease, and its flanking sequences. The DNA containing the amplified protease (or fragment of the gene) is dot-blotted, using standard methods (see *Current Protocols in Molecular Biology, supra*), and the blot is contacted with the oligonucleotide probe. The presence of specific hybridization of the probe to the amplified protease is then detected. Specific hybridization of an allele-specific oligonucleotide probe to DNA from the individual is indicative of a polymorphism in the protease, and is therefore

indicative of a disease or a susceptibility to a disease or condition associated with a protease.

An allele-specific primer hybridizes to a site on target DNA overlapping a polymorphism and only primes amplification of an allelic form to which the primer exhibits perfect complementarity. See Gibbs, *Nucleic Acid Res.* 17, 2427-2448 (1989). This primer is used in conjunction with a second primer, which hybridizes at a distal site. Amplification proceeds from the two primers, resulting in a detectable product, which indicates the particular allelic form is present. A control is usually performed with a second pair of primers, one of which shows a single base mismatch at the polymorphic site and the other of which exhibits perfect complementarity to a distal site. The single-base mismatch prevents amplification and no detectable product is formed. The method works best when the mismatch is included in the 3'-most position of the oligonucleotide aligned with the polymorphism because this position is most destabilizing to elongation from the primer (see, *e.g.*, WO 93/22456).

In another embodiment, arrays of oligonucleotide probes that are complementary to target nucleic acid sequence segments from an individual, can be used to identify polymorphisms in a protease. For example, in one embodiment, an oligonucleotide array can be used. Oligonucleotide arrays typically comprise a plurality of different oligonucleotide probes that are coupled to a surface of a substrate in different known locations. These oligonucleotide arrays, also described as "Genechips™," have been generally described in the art, for example, U.S. Pat. No. 5,143,854 and PCT patent publication Nos. WO 90/15070 and 92/10092. These arrays can generally be produced using mechanical synthesis methods or light directed synthesis methods which incorporate a combination of photolithographic methods and solid phase oligonucleotide synthesis methods. See Fodor *et al.*, *Science* 251:767-777 (1991), Pirrung *et al.*, U.S. Pat. No. 5,143,854 (see also PCT Application No. WO 90/15070) and Fodor *et al.*, PCT Publication No. WO 92/10092 and U.S. Pat. No. 5,424,186, the entire teachings of each of which are incorporated by reference herein. Techniques for the synthesis of these arrays using mechanical synthesis methods are described in, *e.g.*, U.S. Pat. Nos. 5,384,261, the

entire teachings of which are incorporated by reference herein. In another example, linear arrays can be utilized.

Once an oligonucleotide array is prepared, a nucleic acid of interest is hybridized with the array and scanned for polymorphisms. Hybridization and scanning are generally carried out by methods described herein and also in, *e.g.*, published PCT Application Nos. WO 92/10092 and WO 95/11995, and U.S. Pat. No. 5,424,186, the entire teachings of which are incorporated by reference herein. In brief, a target nucleic acid sequence which includes one or more previously identified polymorphic markers is amplified by well known amplification techniques, *e.g.*, PCR. Typically, this involves the use of primer sequences that are complementary to the two strands of the target sequence both upstream and downstream from the polymorphism. Asymmetric PCR techniques may also be used. Amplified target, generally incorporating a label, is then hybridized with the array under appropriate conditions. Upon completion of hybridization and washing of the array, the array is scanned to determine the position on the array to which the target sequence hybridizes. The hybridization data obtained from the scan is typically in the form of fluorescence intensities as a function of location on the array.

Although primarily described in terms of a single detection block, *e.g.*, for detection of a single polymorphism, arrays can include multiple detection blocks, and thus be capable of analyzing multiple, specific polymorphisms. In alternate arrangements, it will generally be understood that detection blocks may be grouped within a single array or in multiple, separate arrays so that varying, optimal conditions may be used during the hybridization of the target to the array. For example, it may often be desirable to provide for the detection of those polymorphisms that fall within G-C rich stretches of a genomic sequence, separately from those falling in A-T rich segments. This allows for the separate optimization of hybridization conditions for each situation.

Additional description of use of oligonucleotide arrays for detection of polymorphisms can be found, for example, in U.S. Patents 5,858,659 and 5,837,832, the entire teachings of which are incorporated by reference herein.

Other methods of nucleic acid analysis can be used to detect polymorphisms

in a protease or variants encoding by a protease. Representative methods include direct manual sequencing (Church and Gilbert, *Proc. Natl. Acad. Sci. USA* 81:1991-1995 (1988); Sanger, F. *et al.*, *Proc. Natl. Acad. Sci. USA* 74:5463-5467 (1977); Beavis *et al.* U.S. Pat. No. 5,288,644); automated fluorescent sequencing; single-stranded conformation polymorphism assays (SSCP); clamped denaturing gel electrophoresis (CDGE); denaturing gradient gel electrophoresis (DGGE) (Sheffield, V.C. *et al.*, *Proc. Natl. Acad. Sci. USA* 86:232-236 (1989)), mobility shift analysis (Orita, M. *et al.*, *Proc. Natl. Acad. Sci. USA* 86:2766-2770 (1989)), restriction enzyme analysis (Flavell *et al.*, *Cell* 15:25 (1978); Geever, *et al.*, *Proc. Natl. Acad. Sci. USA* 78:5081 (1981)); heteroduplex analysis; chemical mismatch cleavage (CMC) (Cotton *et al.*, *Proc. Natl. Acad. Sci. USA* 85:4397-4401 (1985)); RNase protection assays (Myers, R.M. *et al.*, *Science* 230:1242 (1985)); use of polypeptides which recognize nucleotide mismatches, such as *E. coli* mutS protein; allele-specific PCR, for example.

In one embodiment of the invention, diagnosis of a disease or condition associated with a nucleic acid or a susceptibility to a disease or condition associated with a nucleic acid can also be made by expression analysis by quantitative PCR (kinetic thermal cycling). This technique utilizing TaqMan® can be used to allow the identification of polymorphisms and whether a patient is homozygous or heterozygous. The technique can assess the presence of an alteration in the expression or composition of the polypeptide encoded by the nucleic acid or splicing variants encoded by the nucleic acid. Further, the expression of the variants can be quantified as physically or functionally different.

In another embodiment of the invention, diagnosis of a disease or a susceptibility to a disease or condition associated with a protease can also be made by examining expression and/or composition of a protease polypeptide, by a variety of methods, including enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence. A test sample from an individual is assessed for the presence of an alteration in the expression and/or an alteration in composition of the polypeptide encoded by a protease, or for the presence of a particular variant encoded by a protease. An alteration in expression

of a polypeptide encoded by a protease can be, for example, an alteration in the quantitative polypeptide expression (*i.e.*, the amount of polypeptide produced); an alteration in the composition of a polypeptide encoded by a protease is an alteration in the qualitative polypeptide expression (*e.g.*, expression of a mutant protease polypeptide or of a different splicing variant). In a preferred embodiment, diagnosis of a disease or a susceptibility to a disease or condition associated with a protease is made by detecting a particular splicing variant encoded by that protease, or a particular pattern of splicing variants.

Both such alterations (quantitative and qualitative) can also be present. An "alteration" in the polypeptide expression or composition, as used herein, refers to an alteration in expression or composition in a test sample, as compared with the expression or composition of polypeptide by a protease in a control sample. A control sample is a sample that corresponds to the test sample (*e.g.*, is from the same type of cells), and is from an individual who is not affected by a susceptibility to a disease or condition associated with a protease. An alteration in the expression or composition of the polypeptide in the test sample, as compared with the control sample, is indicative of a disease or a susceptibility to a disease or condition associated with a protease. Similarly, the presence of one or more different splicing variants in the test sample, or the presence of significantly different amounts of different splicing variants in the test sample, as compared with the control sample, is indicative of a disease or a susceptibility to a disease or condition associated with a protease. Various means of examining expression or composition of the polypeptide encoded by a protease can be used, including spectroscopy, colorimetry, electrophoresis, isoelectric focusing, and immunoassays (*e.g.*, David *et al.*, U.S. Pat. No. 4,376,110) such as immunoblotting (see also *Current Protocols in Molecular Biology*, particularly Chapter 10). For example, in one embodiment, an antibody capable of binding to the polypeptide (*e.g.*, as described above), preferably an antibody with a detectable label, can be used. Antibodies can be polyclonal, or more preferably, monoclonal. An intact antibody, or a fragment thereof (*e.g.*, Fab or $F(ab')_2$) can be used. The term "labeled", with regard to the probe or antibody, is intended to encompass direct labeling of the probe or antibody by coupling (*i.e.*,

physically linking) a detectable substance to the probe or antibody, as well as indirect labeling of the probe or antibody by reactivity with another reagent that is directly labeled. Examples of indirect labeling include detection of a primary antibody using a fluorescently labeled secondary antibody and end-labeling of a DNA probe with biotin such that it can be detected with fluorescently labeled streptavidin.

Western blotting analysis, using an antibody as described above that specifically binds to a polypeptide encoded by a mutant protease, or an antibody that specifically binds to a polypeptide encoded by a non-mutant gene, or an antibody that specifically binds to a particular splicing variant encoded by a protease, can be used to identify the presence in a test sample of a particular splicing variant or of a polypeptide encoded by a polymorphic or mutant protease, or the absence in a test sample of a particular splicing variant or of a polypeptide encoded by a non-polymorphic or non-mutant gene. The presence of a polypeptide encoded by a polymorphic or mutant gene, or the absence of a polypeptide encoded by a non-polymorphic or non-mutant gene, is diagnostic for a disease or a susceptibility to a disease or condition associated with a protease, as is the presence (or absence) of particular splicing variants encoded by the protease gene.

In one embodiment of this method, the level or amount of polypeptide encoded by a protease in a test sample is compared with the level or amount of the polypeptide encoded by the protease in a control sample. A level or amount of the polypeptide in the test sample that is higher or lower than the level or amount of the polypeptide in the control sample, such that the difference is statistically significant, is indicative of an alteration in the expression of the polypeptide encoded by the protease, and is diagnostic for a disease or a susceptibility to a disease or condition associated with that protease. Alternatively, the composition of the polypeptide encoded by a protease in a test sample is compared with the composition of the polypeptide encoded by the protease in a control sample (*e.g.*, the presence of different splicing variants). A difference in the composition of the polypeptide in the test sample, as compared with the composition of the polypeptide in the control sample, is diagnostic for a disease or a susceptibility to a disease or condition

associated with that protease. In another embodiment, both the level or amount and the composition of the polypeptide can be assessed in the test sample and in the control sample. A difference in the amount or level of the polypeptide in the test sample, compared to the control sample; a difference in composition in the test sample, compared to the control sample; or both a difference in the amount or level, and a difference in the composition, is indicative of a disease or a susceptibility to a disease or condition associated with that protease.

The invention further pertains to a method for the diagnosis and identification of susceptibility to disease in an individual, by identifying an at-risk haplotype in a protease gene. In one embodiment, the at-risk haplotype is one which confers a significant risk of a disease associated with a protease. In one embodiment, significance associated with a haplotype is measured by an odds ratio. In a further embodiment, the significance is measured by a percentage. In one embodiment, a significant risk is measured as an odds ratio of at least about 2.2, including by not limited to: 1.2, 1.3, 1.4, 1.5, 1.6, 1.7, 1.8, and 1.9. In a further embodiment, an odds ratio of at least 1.2 is significant. In a further embodiment, an odds ratio of at least about 1.5 is significant. In a further embodiment, a significant increase in risk is at least about 1.7 is significant. In a further embodiment, a significant increase in risk is at least about 20%, including but not limited to about 25%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, and 98%. In a further embodiment, a significant increase in risk is at least about 50%. It is understood however, that identifying whether a risk is medically significant may also depend on a variety of factors, including the specific disease, the haplotype, and often, environmental factors.

The invention also pertains to methods of diagnosing a disease or a susceptibility to a disease associated with a protease in an individual, comprising screening for an at-risk haplotype in the nucleic acid that is more frequently present in an individual susceptible to a protease associated disease (affected), compared to the frequency of its presence in a healthy individual (control), wherein the presence of the haplotype is indicative of a disease or a susceptibility to a protease associated disease. Standard techniques for genotyping for the presence of SNPs and/or

microsatellite markers that are associated with a protease gene can be used, such as fluorescent based techniques (Chen, *et al.*, *Genome Res. 9*, 492 (1999)), PCR, LCR, Nested PCR and other techniques for nucleic acid amplification. In a certain embodiment, the method comprises assessing in an individual the presence or frequency of SNPs and/or microsatellites in the nucleic acid that are associated with a protease associated disease, wherein an excess or higher frequency of the SNPs and/or microsatellites compared to a healthy control individual is indicative that the individual has a protease associated disease or is susceptible to a protease associated disease.

Kits (*e.g.*, reagent kits) useful in the methods of diagnosis comprise components useful in any of the methods described herein, including for example, hybridization probes or primers as decribed herein (*e.g.*, labeled probes or primers), reagents for detection of labeled molecules, restriction enzymes (*e.g.*, for RFLP analysis), allele-specific oligonucleotides, antibodies which bind to altered or to non-altered (native) protease polypeptide, means for amplification of nucleic acids comprising a protease, or means for analyzing the nucleic acid sequence of a protease or for analyzing the amino acid sequence of a protease polypeptide, etc.

In one embodiment, a kit for diagnosing a protease associated disease or susceptibility to a protease associated disease can comprise primers for nucleic acid amplification of a region in a nucleic acid comprising an at-risk haplotype that is more frequently present in an individual having a protease associated disease or is susceptible to a protease associated disease. The primers can be designed using protion of the nucleic acids flanking SNPs that are indicative of a protease associated disease.

SCREENING ASSAYS AND AGENTS IDENTIFIED THEREBY

The invention provides methods (also referred to herein as "screening assays") for identifying the presence of a nucleotide that hybridizes to a nucleic acid of the invention, as well as for identifying the presence of a polypeptide encoded by a nucleic acid of the invention. In one embodiment, the presence (or absence) of a nucleic acid molecule of interest (*e.g.*, a nucleic acid that has significant homology

with a nucleic acid of the invention) in a sample can be assessed by contacting the sample with a nucleic acid comprising a nucleic acid of the invention (*e.g.*, a nucleic acid having the sequence of one of SEQ ID NOs:1-47, or the complement thereof, or a nucleic acid encoding an amino acid having the sequence of one of SEQ ID NOs:48-94, or a fragment or variant of such nucleic acids), under stringent conditions as described above, and then assessing the sample for the presence (or absence) of hybridization. In a preferred embodiment, high stringency conditions are conditions appropriate for selective hybridization. In another embodiment, a sample containing the nucleic acid molecule of interest is contacted with a nucleic acid containing a contiguous nucleotide sequence (*e.g.*, a primer or a probe as described above) that is at least partially complementary to a part of the nucleic acid molecule of interest (*e.g.*, a protease nucleic acid), and the contacted sample is assessed for the presence or absence of hybridization. In a preferred embodiment, the nucleic acid containing a contiguous nucleotide sequence is completely complementary to a part of the nucleic acid molecule of interest.

In any of these embodiments, all or a portion of the nucleic acid of interest can be subjected to amplification prior to performing the hybridization.

In another embodiment, the presence (or absence) of a polypeptide of interest, such as a polypeptide of the invention or a fragment or variant thereof, in a sample can be assessed by contacting the sample with an antibody that specifically hybridizes to the polypeptide of interest (*e.g.*, an antibody such as those described above), and then assessing the sample for the presence (or absence) of binding of the antibody to the polypeptide of interest.

In another embodiment, the invention provides methods for identifying agents (*e.g.*, fusion proteins, polypeptides, peptidomimetics, prodrugs, receptors, binding agents, antibodies, small molecules or other drugs, or ribozymes which alter (*e.g.*, increase or decrease) the activity of the polypeptides described herein, or which otherwise interact with the polypeptides herein. For example, such agents can be agents which bind to polypeptides described herein (*e.g.*, protease binding agents); which have a stimulatory or inhibitory effect on, for example, activity of polypeptides of the invention; or which change (*e.g.*, enhance or inhibit) the ability

of the polypeptides of the invention to interact with protease binding agents (*e.g.*, receptors or other binding agents); or which alter posttranslational processing of the protease polypeptide (*e.g.*, agents that alter proteolytic processing to direct the polypeptide from where it is normally synthesized to another location in the cell, such as the cell surface; agents that alter proteolytic processing such that more polypeptide is released from the cell, etc.

In one embodiment, the invention provides assays for screening candidate or test agents that bind to or modulate the activity of polypeptides described herein (or biologically active portion(s) thereof), as well as agents identifiable by the assays. Test agents can be obtained using any of the numerous approaches in combinatorial library methods known in the art, including: biological libraries; spatially addressable parallel solid phase or solution phase libraries; synthetic library methods requiring deconvolution; the 'one-bead one-compound' library method; and synthetic library methods using affinity chromatography selection. The biological library approach is limited to polypeptide libraries, while the other four approaches are applicable to polypeptide, non-peptide oligomer or small molecule libraries of compounds (Lam, K.S., *Anticancer Drug Des.* 12:145 (1997)).

In one embodiment, to identify agents which alter the activity of a protease polypeptide, a cell, cell lysate, or solution containing or expressing a protease polypeptide (*e.g.*, one of SEQ ID NOs:48-94, or another splicing variant encoded by a protease), or a fragment or derivative thereof (as described above), can be contacted with an agent to be tested; alternatively, the polypeptide can be contacted directly with the agent to be tested. The level (amount) of protease activity is assessed (*e.g.*, the level (amount) of protease activity is measured, either directly or indirectly), and is compared with the level of activity in a control (*i.e.*, the level of activity of the protease polypeptide or active fragment or derivative thereof in the absence of the agent to be tested). If the level of the activity in the presence of the agent differs, by an amount that is statistically significant, from the level of the activity in the absence of the agent, then the agent is an agent that alters the activity of a protease polypeptide. An increase in the level of protease activity relative to a control, indicates that the agent is an agent that enhances (is an agonist of) protease

activity. Similarly, a decrease in the level of protease activity relative to a control, indicates that the agent is an agent that inhibits (is an antagonist of) protease activity. In another embodiment, the level of activity of a protease polypeptide or derivative or fragment thereof in the presence of the agent to be tested, is compared with a control level that has previously been established. A level of the activity in the presence of the agent that differs from the control level by an amount that is statistically significant indicates that the agent alters protease activity.

The present invention also relates to an assay for identifying agents which alter the expression of a protease gene (*e.g.*, antisense nucleic acids, fusion proteins, polypeptides, peptidomimetics, prodrugs, receptors, binding agents, antibodies, small molecules or other drugs, or ribozymes) which alter (*e.g.*, increase or decrease) expression (*e.g.*, transcription or translation) of the gene or which otherwise interact with the nucleic acids described herein, as well as agents identifiable by the assays. For example, a solution containing a nucleic acid encoding a protease polypeptide (*e.g.*, a protease gene) can be contacted with an agent to be tested. The solution can comprise, for example, cells containing the nucleic acid or cell lysate containing the nucleic acid; alternatively, the solution can be another solution which comprises elements necessary for transcription/translation of the nucleic acid. Cells not suspended in solution can also be employed, if desired. The level and/or pattern of protease expression (*e.g.*, the level and/or pattern of mRNA or of protein expressed, such as the level and/or pattern of different splicing variants) is assessed, and is compared with the level and/or pattern of expression in a control (*i.e.*, the level and/or pattern of the protease expression in the absence of the agent to be tested). If the level and/or pattern in the presence of the agent differs, by an amount or in a manner that is statistically significant, from the level and/or pattern in the absence of the agent, then the agent is an agent that alters the expression of a protease. Enhancement of protease expression indicates that the agent is an agonist of protease activity. Similarly, inhibition of protease expression indicates that the agent is an antagonist of protease activity. In another embodiment, the level and/or pattern of protease polypeptide(s) (*e.g.*, different splicing variants) in the presence of the agent to be tested, is compared with a control level and/or pattern that has previously been

established. A level and/or pattern in the presence of the agent that differs from the control level and/or pattern by an amount or in a manner that is statistically significant indicates that the agent alters protease expression. Such methods can be used to identify compounds that inhibit proteases involved in infection by pathogens.

In another embodiment of the invention, agents which alter the expression of a protease gene or which otherwise interact with the nucleic acids described herein, can be identified using a cell, cell lysate, or solution containing a nucleic acid encoding the promoter region of the protease gene operably linked to a reporter gene. After contact with an agent to be tested, the level of expression of the reporter gene (*e.g.*, the level of mRNA or of protein expressed) is assessed, and is compared with the level of expression in a control (*i.e.*, the level of the expression of the reporter gene in the absence of the agent to be tested). If the level in the presence of the agent differs, by an amount or in a manner that is statistically significant, from the level in the absence of the agent, then the agent is an agent that alters the expression of the protease, as indicated by its ability to alter expression of a gene that is operably linked to the protease gene promoter. Enhancement of the expression of the reporter indicates that the agent is an agonist of protease activity. Similarly, inhibition of the expression of the reporter indicates that the agent is an antagonist of protease activity. In another embodiment, the level of expression of the reporter in the presence of the agent to be tested, is compared with a control level that has previously been established. A level in the presence of the agent that differs from the control level by an amount or in a manner that is statistically significant indicates that the agent alters expression.

Agents which alter the amounts of different splicing variants encoded by a protease (*e.g.*, an agent which enhances activity of a first splicing variant, and which inhibits activity of a second splicing variant), as well as agents which are agonists of activity of a first splicing variant and antagonists of activity of a second splicing variant, can easily be identified using these methods described above.

In other embodiments of the invention, assays can be used to assess the impact of a test agent on the activity of a polypeptide in relation to a protease binding agent. For example, a cell that expresses a compound that interacts with a

protease (herein referred to as a "protease binding agent", which can be a polypeptide or other molecule that interacts with a protease, such as a receptor) is contacted with a protease in the presence of a test agent, and the ability of the test agent to alter the interaction between the protease and the protease binding agent is determined. Alternatively, a cell lysate or a solution containing the protease binding agent, can be used. An agent which binds to the protease or the protease binding agent can alter the interaction by interfering with, or enhancing the ability of the protease to bind to, associate with, or otherwise interact with the protease binding agent. Determining the ability of the test agent to bind to a protease or a protease binding agent can be accomplished, for example, by coupling the test agent with a radioisotope or enzymatic label such that binding of the test agent to the polypeptide can be determined by detecting the labeled with ^{125}I, ^{35}S, ^{14}C or ^{3}H, either directly or indirectly, and the radioisotope detected by direct counting of radioemmission or by scintillation counting. Alternatively, test agents can be enzymatically labeled with, for example, horseradish peroxidase, alkaline phosphatase, or luciferase, and the enzymatic label detected by determination of conversion of an appropriate substrate to product. It is also within the scope of this invention to determine the ability of a test agent to interact with the polypeptide without the labeling of any of the interactants. For example, a microphysiometer can be used to detect the interaction of a test agent with a protease or a protease binding agent without the labeling of either the test agent, protease, or the protease binding agent. McConnell, H.M. *et al.*, *Science* 257:1906-1912 (1992). As used herein, a "microphysiometer" (*e.g.*, Cytosensor™) is an analytical instrument that measures the rate at which a cell acidifies its environment using a light-addressable potentiometric sensor (LAPS). Changes in this acidification rate can be used as an indicator of the interaction between ligand and polypeptide. Thus, these receptors can be used to screen for compounds that are agonists for use in treating a susceptibility to a disease or condition associated with a protease or antagonists for studying a susceptibility to a disease or condition associated with a protease. Drugs could be designed to regulate protease activation which in turn can be used to regulate signaling pathways and transcription events of genes downstream.

In another embodiment of the invention, assays can be used to identify polypeptides that interact with one or more protease polypeptides, as described herein. For example, a yeast two-hybrid system such as that described by Fields and Song (Fields, S. and Song, O., *Nature* 340:245-246 (1989)) can be used to identify polypeptides that interact with one or more protease polypeptides. In such a yeast two-hybrid system, vectors are constructed based on the flexibility of a transcription factor which has two functional domains (a DNA binding domain and a transcription activation domain). If the two domains are separated but fused to two different proteins that interact with one another, transcriptional activation can be achieved, and transcription of specific markers (*e.g.*, nutritional markers such as His and Ade, or color markers such as lacZ) can be used to identify the presence of interaction and transcriptional activation. For example, in the methods of the invention, a first vector is used which includes a nucleic acid encoding a DNA binding domain and also a protease polypeptide, splicing variant, or fragment or derivative thereof, and a second vector is used which includes a nucleic acid encoding a transcription activation domain and also a nucleic acid encoding a polypeptide which potentially may interact with the protease polypeptide, splicing variant, or fragment or derivative thereof (*e.g.*, a protease polypeptide binding agent or receptor). Incubation of yeast containing the first vector and the second vector under appropriate conditions (*e.g.*, mating conditions such as used in the Matchmaker™ system from Clontech (Palo Alto, California, USA)) allows identification of colonies which express the markers of interest. These colonies can be examined to identify the polypeptide(s) which interact with the protease polypeptide or fragment or derivative thereof. Such polypeptides may be useful as agents which alter the activity of expression of a protease polypeptide, as described above.

In more than one embodiment of the above assay methods of the present invention, it may be desirable to immobilize either the protease, the protease binding agent, or other components of the assay on a solid support, in order to facilitate separation of complexed from uncomplexed forms of one or both of the polypeptides, as well as to accommodate automation of the assay. Binding of a test agent to the polypeptide, or interaction of the polypeptide with a binding agent in the

presence and absence of a test agent, can be accomplished in any vessel suitable for containing the reactants. Examples of such vessels include microtitre plates, test tubes, and micro-centrifuge tubes. In one embodiment, a fusion protein (*e.g.*, a glutathione-S-transferase fusion protein) can be provided which adds a domain that allows a protease or a protease binding agent to be bound to a matrix or other solid support.

In another embodiment, modulators of expression of nucleic acid molecules of the invention are identified in a method wherein a cell, cell lysate, or solution containing a nucleic acid encoding a protease is contacted with a test agent and the expression of appropriate mRNA or polypeptide (*e.g.*, splicing variant(s)) in the cell, cell lysate, or solution, is determined. The level of expression of appropriate mRNA or polypeptide(s) in the presence of the test agent is compared to the level of expression of mRNA or polypeptide(s) in the absence of the test agent. The test agent can then be identified as a modulator of expression based on this comparison. For example, when expression of mRNA or polypeptide is greater (statistically significantly greater) in the presence of the test agent than in its absence, the test agent is identified as a stimulator or enhancer of the mRNA or polypeptide expression. Alternatively, when expression of the mRNA or polypeptide is less (statistically significantly less) in the presence of the test agent than in its absence, the test agent is identified as an inhibitor of the mRNA or polypeptide expression. The level of mRNA or polypeptide expression in the cells can be determined by methods described herein for detecting mRNA or polypeptide.

This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein in an appropriate animal model. For example, an agent identified as described herein (*e.g.*, a test agent that is a modulating agent, an antisense nucleic acid molecule, a specific antibody, or a polypeptide-binding agent) can be used in an animal model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal model to determine the mechanism of action of such an agent. Furthermore, this invention

pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein. In addition, an agent identified as described herein can be used to alter activity of a polypeptide encoded by a protease, or to alter expression of a protease, by contacting the polypeptide or the gene (or contacting a cell comprising the polypeptide or the gene) with the agent identified as described herein.

PHARMACEUTICAL COMPOSITIONS

The present invention also pertains to pharmaceutical compositions comprising nucleic acids described herein, particularly nucleotides encoding the polypeptides described herein; comprising polypeptides described herein (*e.g.*, one or more of SEQ ID NOs:48-94); and/or comprising other splicing variants encoded by a protease; and/or an agent that alters (*e.g.*, enhances or inhibits) protease gene expression or protease polypeptide activity as described herein. For instance, a polypeptide, protein (*e.g.*, a protease receptor), an agent that alters protease gene expression, or a protease binding agent or binding partner, fragment, fusion protein or prodrug thereof, or a nucleotide or nucleic acid construct (vector) comprising a nucleotide of the present invention, or an agent that alters protease polypeptide activity, can be formulated with a physiologically acceptable carrier or excipient to prepare a pharmaceutical composition. The carrier and composition can be sterile. The formulation should suit the mode of administration.

Suitable pharmaceutically acceptable carriers include but are not limited to water, salt solutions (*e.g.*, NaCl), saline, buffered saline, alcohols, glycerol, ethanol, gum arabic, vegetable oils, benzyl alcohols, polyethylene glycols, gelatin, carbohydrates such as lactose, amylose or starch, dextrose, magnesium stearate, talc, silicic acid, viscous paraffin, perfume oil, fatty acid esters, hydroxymethylcellulose, polyvinyl pyrolidone, etc., as well as combinations thereof. The pharmaceutical preparations can, if desired, be mixed with auxiliary agents, *e.g.*, lubricants, preservatives, stabilizers, wetting agents, emulsifiers, salts for influencing osmotic pressure, buffers, coloring, flavoring and/or aromatic substances and the like which do not deleteriously react with the active agents.

The composition, if desired, can also contain minor amounts of wetting or emulsifying agents, or pH buffering agents. The composition can be a liquid solution, suspension, emulsion, tablet, pill, capsule, sustained release formulation, or powder. The composition can be formulated as a suppository, with traditional binders and carriers such as triglycerides. Oral formulation can include standard carriers such as pharmaceutical grades of mannitol, lactose, starch, magnesium stearate, polyvinyl pyrollidone, sodium saccharine, cellulose, magnesium carbonate, etc.

Methods of introduction of these compositions include, but are not limited to, intradermal, intramuscular, intraperitoneal, intraocular, intravenous, subcutaneous, topical, oral and intranasal. Other suitable methods of introduction can also include gene therapy (as described below), rechargeable or biodegradable devices, particle acceleration devises ("gene guns") and slow release polymeric devices. The pharmaceutical compositions of this invention can also be administered as part of a combinatorial therapy with other agents.

The composition can be formulated in accordance with the routine procedures as a pharmaceutical composition adapted for administration to human beings. For example, compositions for intravenous administration typically are solutions in sterile isotonic aqueous buffer. Where necessary, the composition may also include a solubilizing agent and a local anesthetic to ease pain at the site of the injection. Generally, the ingredients are supplied either separately or mixed together in unit dosage form, for example, as a dry lyophilized powder or water free concentrate in a hermetically sealed container such as an ampule or sachette indicating the quantity of active agent. Where the composition is to be administered by infusion, it can be dispensed with an infusion bottle containing sterile pharmaceutical grade water, saline or dextrose/water. Where the composition is administered by injection, an ampule of sterile water for injection or saline can be provided so that the ingredients may be mixed prior to administration.

For topical application, nonsprayable forms, viscous to semi-solid or solid forms comprising a carrier compatible with topical application and having a dynamic viscosity preferably greater than water, can be employed. Suitable formulations

include but are not limited to solutions, suspensions, emulsions, creams, ointments, powders, enemas, lotions, sols, liniments, salves, aerosols, etc., which are, if desired, sterilized or mixed with auxiliary agents, *e.g.*, preservatives, stabilizers, wetting agents, buffers or salts for influencing osmotic pressure, etc. The agent may be incorporated into a cosmetic formulation. For topical application, also suitable are sprayable aerosol preparations wherein the active ingredient, preferably in combination with a solid or liquid inert carrier material, is packaged in a squeeze bottle or in admixture with a pressurized volatile, normally gaseous propellant, *e.g.*, pressurized air.

Agents described herein can be formulated as neutral or salt forms. Pharmaceutically acceptable salts include those formed with free amino groups such as those derived from hydrochloric, phosphoric, acetic, oxalic, tartaric acids, etc., and those formed with free carboxyl groups such as those derived from sodium, potassium, ammonium, calcium, ferric hydroxides, isopropylamine, triethylamine, 2-ethylamino ethanol, histidine, procaine, etc.

The agents are administered in a therapeutically effective amount. The amount of agents which will be therapeutically effective in the treatment of a particular disorder or condition will depend on the nature of the disorder or condition, and can be determined by standard clinical techniques. In addition, *in vitro* or *in vivo* assays may optionally be employed to help identify optimal dosage ranges. The precise dose to be employed in the formulation will also depend on the route of administration, and the seriousness of the symptoms of a susceptibility to a disease or condition associated with a protease, and should be decided according to the judgment of a practitioner and each patient's circumstances. Effective doses may be extrapolated from dose-response curves derived from *in vitro* or animal model test systems.

The invention also provides a pharmaceutical pack or kit comprising one or more containers filled with one or more of the ingredients of the pharmaceutical compositions of the invention. Optionally associated with such container(s) can be a notice in the form prescribed by a governmental agency regulating the manufacture, use or sale of pharmaceuticals or biological products, which notice reflects approval

by the agency of manufacture, use of sale for human administration. The pack or kit can be labeled with information regarding mode of administration, sequence of drug administration (*e.g.*, separately, sequentially or concurrently), or the like. The pack or kit may also include means for reminding the patient to take the therapy. The pack or kit can be a single unit dosage of the combination therapy or it can be a plurality of unit dosages. In particular, the agents can be separated, mixed together in any combination, present in a single vial or tablet. Agents assembled in a blister pack or other dispensing means is preferred. For the purpose of this invention, unit dosage is intended to mean a dosage that is dependent on the individual pharmacodynamics of each agent and administered in FDA approved dosages in standard time courses.

METHODS OF THERAPY

The present invention also pertains to methods of treatment (prophylactic and/or therapeutic) for a disease or condition associated with a protease or susceptibility to a disease or condition associated with a protease, using a protease therapeutic agent. A "protease therapeutic agent" is an agent that alters (*e.g.*, enhances or inhibits) protease polypeptide activity and/or protease gene expression, as described herein (*e.g.*, a protease agonist or antagonist). Protease therapeutic agents can alter protease polypeptide activity or gene expression by a variety of means, such as, for example, by providing additional protease polypeptide or by upregulating the transcription or translation of the protease gene; by altering posttranslational processing of the protease polypeptide; by altering transcription of protease splicing variants; or by interfering with protease polypeptide activity (*e.g.*, by binding to a protease polypeptide), or by downregulating the transcription or translation of a protease gene. Representative protease therapeutic agents include the following:

nucleic acids or fragments or derivatives thereof described herein, particularly nucleotides encoding the polypeptides described herein and vectors comprising such nucleic acids (*e.g.*, a gene, cDNA, and/or mRNA, such as a nucleic acid encoding a protease polypeptide or active fragment or derivative thereof, or an

oligonucleotide; for example, one of SEQ ID NOs:1-47, or a complement thereof, or a nucleic acid encoding one of SEQ ID NOs:48-94, or fragments or derivatives thereof);

polypeptides described herein (*e.g.*, one or more of SEQ ID NOs:48-94, and/or other splicing variants encoded by a protease, or fragments or derivatives thereof);

other polypeptides (*e.g.*, protease receptors); protease binding agents; peptidomimetics; fusion proteins or prodrugs thereof; antibodies (*e.g.*, an antibody to a mutant protease polypeptide, or an antibody to a non-mutant protease polypeptide, or an antibody to a particular splicing variant encoded by a protease, as described above); ribozymes; other small molecules; and

other agents that alter (*e.g.*, enhance or inhibit) protease gene expression or polypeptide activity, or that regulate transcription of protease splicing variants (*e.g.*, agents that affect which splicing variants are expressed, or that affect the amount of each splicing variant that is expressed.

More than one protease therapeutic agent can be used concurrently, if desired.

A protease therapeutic agent that is a nucleic acid is used in the treatment of a susceptibility to a disease or condition associated with a protease. The term, "treatment" as used herein, refers not only to ameliorating symptoms associated with the disease, but also preventing or delaying the onset of the disease, and also lessening the severity or frequency of symptoms of the disease. The therapy is designed to alter (*e.g.*, inhibit or enhance), replace or supplement activity of a protease polypeptide in an individual. For example, a protease therapeutic agent can be administered in order to upregulate or increase the expression or availability of the protease gene or of specific splicing variants of protease, or, conversely, to downregulate or decrease the expression or availability of the protease gene or specific splicing variants of the protease. Upregulation or increasing expression or availability of a native protease gene or of a particular splicing variant could interfere with or compensate for the expression or activity of a defective gene or another splicing variant; downregulation or decreasing expression or availability of a

native protease gene or of a particular splicing variant could minimize the expression or activity of a defective gene or the particular splicing variant and thereby minimize the impact of the defective gene or the particular splicing variant.

The protease therapeutic agent(s) are administered in a therapeutically effective amount (*i.e.*, an amount that is sufficient to treat the disease, such as by ameliorating symptoms associated with the disease, preventing or delaying the onset of the disease, and/or also lessening the severity or frequency of symptoms of the disease). The amount which will be therapeutically effective in the treatment of a particular individual's disorder or condition will depend on the symptoms and severity of the disease, and can be determined by standard clinical techniques. In addition, *in vitro* or *in vivo* assays may optionally be employed to help identify optimal dosage ranges. The precise dose to be employed in the formulation will also depend on the route of administration, and the seriousness of the disease or disorder, and should be decided according to the judgment of a practitioner and each patient's circumstances. Effective doses may be extrapolated from dose-response curves derived from *in vitro* or animal model test systems.

In one embodiment, a nucleic acid of the invention (*e.g.*, a nucleic acid encoding a protease polypeptide, such as one of SEQ ID NOs:1-47, or a complement thereof; or another nucleic acid that encodes a protease polypeptide or a splicing variant, derivative or fragment thereof, such as a nucleic acid encoding one of SEQ ID NOs:48-94) can be used, either alone or in a pharmaceutical composition as described above. For example, a protease or a cDNA encoding a protease polypeptide, either by itself or included within a vector, can be introduced into cells (either *in vitro* or *in vivo*) such that the cells produce native protease polypeptide. If necessary, cells that have been transformed with the gene or cDNA or a vector comprising the gene or cDNA can be introduced (or re-introduced) into an individual affected with the disease. Thus, cells which, in nature, lack native protease expression and activity, or have mutant protease expression and activity, or have expression of a disease-associated protease splicing variant, can be engineered to express the protease polypeptide or an active fragment of the protease polypeptide (or a different variant of the protease polypeptide). In a preferred embodiment,

nucleic acid encoding a protease polypeptide, or an active fragment or derivative thereof, can be introduced into an expression vector, such as a viral vector, and the vector can be introduced into appropriate cells in an animal. Other gene transfer systems, including viral and nonviral transfer systems, can be used. Alternatively, nonviral gene transfer methods, such as calcium phosphate coprecipitation, mechanical techniques (*e.g.*, microinjection); membrane fusion-mediated transfer via liposomes; or direct DNA uptake, can also be used.

Alternatively, in another embodiment of the invention, a nucleic acid of the invention; a nucleic acid complementary to a nucleic acid of the invention; or a portion of such a nucleic acid (*e.g.*, an oligonucleotide as described below), can be used in "antisense" therapy, in which a nucleic acid (*e.g.*, an oligonucleotide) which specifically hybridizes to the mRNA and/or genomic DNA of a protease is administered or generated *in situ*. The antisense nucleic acid that specifically hybridizes to the mRNA and/or DNA inhibits expression of the protease polypeptide, *e.g.*, by inhibiting translation and/or transcription. Binding of the antisense nucleic acid can be by conventional base pair complementarity, or, for example, in the case of binding to DNA duplexes, through specific interaction in the major groove of the double helix.

An antisense construct of the present invention can be delivered, for example, as an expression plasmid as described above. When the plasmid is transcribed in the cell, it produces RNA which is complementary to a portion of the mRNA and/or DNA which encodes the protease polypeptide. Alternatively, the antisense construct can be an oligonucleotide probe which is generated *ex vivo* and introduced into cells; it then inhibits expression by hybridizing with the mRNA and/or genomic DNA of the protease. In one embodiment, the oligonucleotide probes are modified oligonucleotides which are resistant to endogenous nucleases, *e.g.*, exonucleases and/or endonucleases, thereby rendering them stable *in vivo*. Exemplary nucleic acid molecules for use as antisense oligonucleotides are phosphoramidate, phosphothioate and methylphosphonate analogs of DNA (see also U.S. Pat. Nos. 5,176,996; 5,264,564; and 5,256,775). Additionally, general approaches to constructing oligomers useful in antisense therapy are also described,

for example, by Van der Krol *et al.* (*Biotechniques* 6:958-976 (1988)); and Stein *et al.* (*Cancer Res.* 48:2659-2668 (1988)). With respect to antisense DNA, oligodeoxyribonucleotides derived from the translation initiation site are preferred.

To perform antisense therapy, oligonucleotides (mRNA, cDNA or DNA) are designed that are complementary to mRNA encoding the protease. The antisense oligonucleotides bind to protease mRNA transcripts and prevent translation. Absolute complementarity, although preferred, is not required. A sequence "complementary" to a portion of an RNA, as referred to herein, indicates that a sequence has sufficient complementarity to be able to hybridize with the RNA, forming a stable duplex; in the case of double-stranded antisense nucleic acids, a single strand of the duplex DNA may thus be tested, or triplex formation may be assayed. The ability to hybridize will depend on both the degree of complementarity and the length of the antisense nucleic acid, as described in detail above. Generally, the longer the hybridizing nucleic acid, the more base mismatches with an RNA it may contain and still form a stable duplex (or triplex, as the case may be). One skilled in the art can ascertain a tolerable degree of mismatch by use of standard procedures.

The oligonucleotides used in antisense therapy can be DNA, RNA, or chimeric mixtures or derivatives or modified versions thereof, single-stranded or double-stranded. The oligonucleotides can be modified at the base moiety, sugar moiety, or phosphate backbone, for example, to improve stability of the molecule, hybridization, etc. The oligonucleotides can include other appended groups such as peptides (*e.g.*, for targeting host cell receptors *in vivo*), or agents facilitating transport across the cell membrane (see, *e.g.*, Letsinger *et al.*, *Proc. Natl. Acad. Sci. USA* 86:6553-6556 (1989); Lemaitre *et al.*, *Proc. Natl. Acad. Sci. USA* 84:648-652 (1987); PCT International Publication No. WO 88/09810) or the blood-brain barrier (see, *e.g.*, PCT International Publication No. WO 89/10134), or hybridization-triggered cleavage agents (see, *e.g.*, Krol *et al.*, *BioTechniques* 6:958-976 (1988)) or intercalating agents. (See, *e.g.*, Zon, Pharm. Res. 5:539-549 (1988)). To this end, the oligonucleotide may be conjugated to another molecule

(*e.g.*, a peptide, hybridization triggered cross-linking agent, transport agent, hybridization-triggered cleavage agent).

The antisense molecules are delivered to cells which express protease *in vivo*. A number of methods can be used for delivering antisense DNA or RNA to cells; *e.g.*, antisense molecules can be injected directly into the tissue site, or modified antisense molecules, designed to target the desired cells (*e.g.*, antisense linked to peptides or antibodies that specifically bind receptors or antigens expressed on the target cell surface) can be administered systematically. Alternatively, in a preferred embodiment, a recombinant DNA construct is utilized in which the antisense oligonucleotide is placed under the control of a strong promoter (*e.g.*, pol III or pol II). The use of such a construct to transfect target cells in the patient results in the transcription of sufficient amounts of single stranded RNAs that will form complementary base pairs with the endogenous protease transcripts and thereby prevent translation of the protease mRNA. For example, a vector can be introduced *in vivo* such that it is taken up by a cell and directs the transcription of an antisense RNA. Such a vector can remain episomal or become chromosomally integrated, as long as it can be transcribed to produce the desired antisense RNA. Such vectors can be constructed by recombinant DNA technology methods standard in the art and described above. For example, a plasmid, cosmid, YAC or viral vector can be used to prepare the recombinant DNA construct which can be introduced directly into the tissue site. Alternatively, viral vectors can be used which selectively infect the desired tissue, in which case administration may be accomplished by another route (*e.g.*, systemically).

Endogenous protease expression can also be reduced by inactivating or "knocking out" protease or its promoter using targeted homologous recombination (*e.g.*, see Smithies *et al.*, *Nature* 317:230-234 (1985); Thomas & Capecchi, *Cell* 51:503-512 (1987); Thompson *et al.*, *Cell* 5:313-321 (1989)). For example, a mutant, non-functional protease (or a completely unrelated DNA sequence) flanked by DNA homologous to the endogenous protease (either the coding regions or regulatory regions of protease) can be used, with or without a selectable marker and/or a negative selectable marker, to transfect cells that express the protease *in*

vivo. Insertion of the DNA construct, via targeted homologous recombination, results in inactivation of the protease. The recombinant DNA constructs can be directly administered or targeted to the required site *in vivo* using appropriate vectors, as described above. Alternatively, expression of non-mutant proteases can be increased using a similar method: targeted homologous recombination can be used to insert a DNA construct comprising a non-mutant, functional protease, *e.g.*, a gene having one of SEQ ID NOs:1-47, or the complement thereof, or a portion thereof, in place of a mutant protease in the cell, as described above. In another embodiment, targeted homologous recombination can be used to insert a DNA construct comprising a nucleic acid that encodes a protease polypeptide variant that differs from that present in the cell.

Alternatively, endogenous protease expression can be reduced by targeting deoxyribonucleotide sequences complementary to the regulatory region of a protease (*i.e.*, the protease promoter and/or enhancers) to form triple helical structures that prevent transcription of the protease in target cells in the body. (See generally, Helene, C., *Anticancer Drug Des.*, 6(6):569-84 (1991); Helene, C. *et al.*, *Ann. N.Y. Acad. Sci.* 660:27-36 (1992); and Maher, L. J., *Bioassays* 14(12):807-15 (1992)). Likewise, the antisense constructs described herein, by antagonizing the normal biological activity of one of the protease proteins, can be used in the manipulation of tissue, *e.g.*, tissue differentiation, both *in vivo* and *for ex vivo* tissue cultures. Furthermore, the anti-sense techniques (*e.g.*, microinjection of antisense molecules, or transfection with plasmids whose transcripts are anti-sense with regard to a protease mRNA or gene sequence) can be used to investigate the role of one or protease in developmental events, as well as the normal cellular function of the proteases in adult tissue. Such techniques can be utilized in cell culture, but can also be used in the creation of transgenic animals.

In yet another embodiment of the invention, other protease therapeutic agents as described herein can also be used in the treatment or prevention of a susceptibility to a disease or condition associated with a protease. The therapeutic agents can be delivered in a composition, as described above, or by themselves. They can be administered systemically, or can be targeted to a particular tissue. The therapeutic

agents can be produced by a variety of means, including chemical synthesis; recombinant production; *in vivo* production (*e.g.*, a transgenic animal, such as U.S. Pat. No. 4,873,316 to Meade *et al.*), for example, and can be isolated using standard means such as those described herein.

A combination of any of the above methods of treatment (*e.g.*, administration of non-mutant protease polypeptide in conjunction with antisense therapy targeting mutant protease mRNA; administration of a first splicing variant encoded by a protease in conjunction with antisense therapy targeting a second splicing encoded by a protease), can also be used.

The teachings of all publications cited herein are incorporated herein by reference in their entirety.

While this invention has been particularly shown and described with references to preferred embodiments thereof, it will be understood by those skilled in the art that various changes in form and details may be made therein without departing from the scope of the invention encompassed by the appended claims.

Appendix I

ctgchr11q_1 MOOSE13873 7240429..7240464, 7242698..7242815, 7243685..7243803, 7263311..7263510, 7265776..7265892, 7267026..7267371, 7268489..7268635 LVDEQPLENYLDMEYFGTIGIGTPAQDFTVVFDTGSSNL WVPSVYCSSLACTNHNRFNPEDSSTYQSTSETVSITYGTGSMTGILGYDTVQ VGGISDTNQIFGLSETEPGSFLYYAPFDGILGLAYPSISSSGATPVFDNIWNQG LVSQDLFSVYLSADDKSGSVVIFGGIDSSYYTGSLNWVPVTVEGYWQITVDS ITMNGETIACAEGCQAIVDTGTSLLTGPTSPIANIQSDIGASENSDGDVSPAPT ALFYTQVVGVPGRSDENPSNFSHPHSFQMVVSCSAISSLPDIVFTINGVQYPV PPSAYILQSEGSCISGFQGMNVPTESGELWILGDVFIRQYFTVFDRANNQVGL APVA (SEQ ID NO: 48)

ctggtagatgaacagcccctggagaactacctggatatggagtacttcggcactatcggcatcggaactcctgcccagga tttcaccgtcgtctttgacaccggctcctccaacctgtgggtgccctcagtctactgctccagtcttgcctgcACCaacca caaccgcttcaaccctgaggattcttccacctaccagtccaccagcgagacagtctccatcacctacggcaccggcagca tgacaggcatcctcggatacgacactgtccaggttggaggcatctctgacaccaatcagatcttcggcctgagcgagacg gaacctggctccttcctgtattatgctcccttcgatggcatcctggggctggcctaccccagcatttcctcctccggggccac acccgtctttgacaacatctggaaccagggcctggtttctcaggacctcttctctgtctacctcagcGCCgatgacaaga gtggcagcgtggtgatctttggtggcattgactcttcttactacactggaagtctgaactgggtgcctgttaccgtcgagggtt actggcagatcaccgtggacAGCatcaccatgaacggagagaccatcgcctgtgctgagggctgccaggccattgtt gacaccggcacctctctgctgaccggcccaaccagccccattgccaacatccagagcgacatcggagccagcgagaac tcagatggcgacgtgagtccagccccgactgccctgttctacactcaagtagtgggtgtgccaggcagaagcgacgaaa acccttctaacttttctcaccctcactctttccagatggtggtcagctgctcagccatcagcagcctgcccgacatcgtcttca ccatcaatggagtccagtaccccgtgccacccagtgcctacatcctgcagagcgaggggagctgcatcagtggcttcca gggcatgaacgtccccaccgaatctggagagctttggatcctgggtgatgtcttcatccgccagtactttaccgtcttcgac agggcaaacaaccaggtcggcctggcccctgtggct (SEQ ID NO: 1)

ctgchr11q_1 MOOSE13874 7136792..7136938, 7138056..7138401, 7139535..7139651, 7141003..7141202, 7141878..7141996, 7142867..7142984, 7145221..7145256 LVDEQPLENYLDMEYFGTIGIGTPAQDFTVLFDTGSSNL WVPSVYCSSLACTNHNRFNPEDSSTYQSTSETVSITYGTGSMTGILGYDTVQ VGGISDTNQIFGLSETEPGSFLYYAPFDGILGLAYPSISSSGATPVFDNIWNQG LVSQDLFSVYLSADDKSGSVVIFGGIDSSYYTGSLNWVPVTVEGYWQITVDS ITMNGETIACAEGCQAIVDTGTSLLTGPTSPIANIQSDIGASENSDGDVSPAPT ALFYTQVVGVPGRSDENPSNFSHPHSFQMVVSCSAISSLPDIVFTINGVQYPV PPSAYILQSEGSCISGFQGMNVPTESGELWILGDVFIRQYFTVFDRANNQVGL APVA (SEQ ID NO: 49)

ctggtagatgaacagcccctggagaactacctggatatggagtacttcggcactatcggcatcggaactcctgcccagga tttcactgtcctctttgacaccggctcctccaacctgtgggtgccctcagtctactgctccagtcttgcctgcACCaaccac aaccgcttcaaccctgaggattcttccacctaccagtccaccagcgagacagtctccatcacctacggcaccggcagcat gacaggcatcctcggatacgacactgtccaggttggaggcatctctgacaccaatcagatcttcggcctgagcgagacgg aacctggctccttcctgtattatgctcccttcgatggcatcctggggctggcctaccccagcatttcctcctccggggccaca cccgtctttgacaacatctggaaccagggcctggtttctcaggacctcttctctgtctacctcagcGCCgatgacaagagt ggcagcgtggtgatctttggtggcattgactcttcttactacactggaagtctgaactgggtgcctgttaccgtcgagggtta ctggcagatcaccgtggacAGCatcaccatgaacggagagaccatcgcctgtgctgagggctgccaggccattgttg acaccggcacctctctgctgaccggcccaaccagccccattgccaacatccagagcgacatcggagccagcgagaact cagatggcgacgtgagtccagccccgactgccctgttctacactcaagtagtgggtgtgccaggcagaagcgacgaaaa

ccctctaacttttctcaccctcactctttccagatggtggtcagctgctcagccatcagcagcctgcccgacatcgtcttcac
catcaatggagtccagtaccccgtgccacccagtgcctacatcctgcagagcgaggggagctgcatcagtggcttccag
ggcatgaacgtccccaccgaatctggagagctttggatcctgggtgatgtcttcatccgccagtactttaccgtcttcgaca
gggcaaacaaccaggtcggcctggcccctgtggct (SEQ ID NO: 2)

ctg10008 MOOSE13895 396108..396195, 396236..396353, 396946..397108, 397589..397813, 398026..398172, 398273..398362, 420625..420664, 456670..456713 DKAWLKRGNKQFNEGKESDRCLIFKCKNKDVKMIEQH NQEYSQGKHSFTMAMNAFGDMTNEEFRQVMNGFQYQKHRKGKQFQERLL LEIPTSVDWREKGYMTPVKDQGQCGSCWAFSATGALEGQMFWKTGKLISL NEQNLVDCSGPQGNEGCNGDFMDNPFRYVQENGGLDSEASYPYEGKVKTC RYNPKYSAANDTGFVDIPSREKDLAKAVATVGPISVAVGASHVFFQFYKKGI YFEPRCDPEGLDHAMLVVGYSYEGADSDNNKYWLVKNRESFFAISIYIPQAL NHISEIPEVFFPLRQ (SEQ ID NO: 50)

gacaaggcatggttgaaaagagggaacaaacagttcaatgaaGGAaaggaatctgatagatgtctgatatttaaatgta
aaaataaggacgtgaagatgattgagcagcacaatcaggaatacagccaagggaaacacagcttcacaatggccatgaa
cgcctttggagacatgaccaatgaagaattcaggcaggtgatgaatggttttcaataccagaagcacaggaaggggaaac
agttccaggaacgcctgcttcttgagatccccacatctgtggactggagagagaaaggctacatgactcctgtgaaggatc
agggtcagtgtggctcttgttgggcttttagtgcaactggtgctctggaagggcagatgttctggaaaacaggcaaacttat
ctcactgaatgagcagaatctggtagactgctctgggcctcaaggcaatgagggctgcaatggtgacttcatggataatcc
cttccggtatgttcaggagaacggaggcctggactctgaggcatcctatccatatgaaggaaaggttaaaacctgtaggta
caatcccaagtattctgctgctaatgacactggttttgtggacatcccttcacgggagaaggacctggcgaaggcagtggc
aactgtggggcccatctctgttgctgttggtgcaagccatgtcttcttccagttctataaaaaaGGAatttatttgagccac
gctgtgaccctgaaggcctggatcatgctatgctggtggttggctacagctatgaaggagcagactcagataacaataaat
attggctggtgaagaacAGGgaatccttttttgcaatcagtatttacataccacaagcccttaatcacatttctgaaatccca
gaagtctttttccccccttaggcag (SEQ ID NO: 3)

ctgChr_13ctg3 MOOSE13908 31779459..31779493, 31784161..31784194, 31786584..31786698, 31787274..31787332, 31787822..31787839, 31788078..31788126, 31796102..31796145, 31796615..31796690, 31806469..31806527, 31807429..31807509 SNVKPRKKNPKGFTVNVLANIGVINHWKLPKLTHKGEKNSGKEGEKPHKTN SASNSCGTISNLGSIENGTNEGKYKINFILKVILKNANTVRTEKRNRQIHNYN CLHCTVETSDYQVQAPSIDEKVDLHFIALVHVDGHLYELDGRKPFPINHGET SDETLLEDAIEVCKKFMERDPDELRFNAIALSAA (SEQ ID NO: 51)

tcaaatgtgaagccaagaaagaaaaatccaaaaGGGttcactgttaatgtcctggcaaacattggtgtcataaatcactg
gaaattacccaaacttactcacaaaggggaaaaaaatagtggaaaagaggggggaaaaaccacacaaaacaaacagtgc
atccaacagctgtgggacaatttcaAATttaggaagcattgaaaatggaacaaatgaaggtaaatataaaataaattttatt
cttaaagtcattttaaaaaacgcaaacactgttagaactgaaaagagaaatagacaaattcacaattatAACtgtttacact
gcactgttgaaacatcagattatcaagttcaagcaccaagtatagatgagaaagtagatcttcattttattgcattagttcatgt
agatgggcatctctatgaattaGATgggcggaagccatttccaattaaccatggtgaaactagtgatgaaactttattaga
ggatgccatagaagtttgcaagaagtttatggagcgcgaccctgatgaactaagatttaatgcgattgctctttctgcagca
(SEQ ID NO: 4)

ctg15ctg27 MOOSE13930 800712..800756, 825940..826087, 826815..827041, 828426..828548, 855490..855558 ACLVSGFGPFRQHLVNSSWEAVKELSKLGLGNETVVQLRTLELPVDYREAK RRVTGIWEDHQPQLVVHVGMDTAAKAIILEQSGKNQGYRDADIRSFWPEGG VCLPGSPDVLESGVCMKAVCKRVAVEGVDVIFSRDAGRYVCDYTYYLSLH

HGKGCAALIHVPPLSRGLPASLLGRALRVIIQEMLEEAGEKQKEVTASGTSH
(SEQ ID NO: 52)
gcatgtttggtttcaggttttgggcccttccggcagcacttggtgaattccagctgggaagcagtgaaggagctctccaagc
tgggcctggggaatgaaacagtggtgcagctgcggactctggagctgcctgtagattacagggaggctaagcggaggg
tcaccggaatctgggaagatcatcagccgcaactcgtcgtgcatgtgggcatggacaccgccgccaaggcgatcattctg
gaacagtctggcaagaaccaaggctaccgggacgccgacatccgcagcttctggcccgagggcggcgtgtgcctacct
ggcagcccagacgtgctggagtcaggggtctgcatgaaggcagtctgcaagcgcgtagctgtggagggtgtcgacgtg
atcttttcccgagatgcaggcAGAtacgtctgtgattatacctattacctgtctctgcatcatggaaagggctgcgcggca
ctcatccatgtccctccactatcgcgcgggctcccggccagcctgctgggaagagccttgagagtcatcatccaggaaat
gctggaagaggcaggggagaaacaaaaagaggtgacggcttctggcacatcccac (SEQ ID NO: 5)
ctgChr_Xctg264 MOOSE13940 1447207..1448280
RKKSVYTVGLRGLINLGNTCFMNCIVQALTHIPLLKDFFLSDKHKCIMTSPSL
CLVCEMSSLFHAMYSGSRTPHIPYKLLHLIWIHAEHLAGYRQQDAHEFLIAIL
DVLHRHSKDDSGGQEANNPNCCNCIIDQIFTGGLQSDVTCQACHSVSTTIDP
CWDISLDLPGSCATFDSQNPERADSTVSRDDHIPGIPSLTDCLQWFTRPEHLG
SSAKIKCNSCQSYQESTKQLTMKKLPIVACFHLKRFEHVGKQRRKINTFISFP
LELDMTPFLASTKESRMKEGQPPTDCVPNENKYSLFAVINHHGTLESGHYTS
FIRQQKDQWFSCDDAIITKATIEDLLYSEGYLLFYHKQGLEK (SEQ ID NO: 53)
agaaaaaagtcagtctatactgtaggcctgagagggctaatcaatcttgggaacacttgttttatgaattgtattgt
ccaggcacttacccatattcctctactgaaagatttcttcctctctgacaagcacaaatgtataatgacaagccccagcttgtg
tctggtctgtgaaatgtcttcgctttttcatgctatgtactctgggagccgaactcctcacattccctataagttactgcatctgat
atggatccatgcagaacatttagcagggtacaggcagcaggatgcccatgagttccttattgcaatattagacgtgctacat
agacacagcaaagatgatagtggtgggcaggaggccaataaccccaactgctgtaactgcatcatagaccaaatctttac
aggtggcctgcaatcagatgtcacatgtcaagcctgccatagtgtttctaccaccatagacccatgctgggacatcagtttg
gacttgcctggctcttgtgccacattcgattcccagaacccagagagggctgacagcacagtgagcagggatgaccacat
accaggaatcccctcacttacagactgtctacagtggtttacaaggccagagcacctaggaagcagtgccaaaatcaaat
gcaatagttgccaaagctaccaggagtctactaaacagctcacaatgaaaaaattacccattgtggcttgttttcatctcaag
cggtttgagcatgtaggcaaacagaggcgaaagattaatacctttatctcctttcccttggagctggacatgactccgtttttg
gcctctactaaagagagcagaatgaaagaaggccagccaccaacagattgtgtgcccaatgagaataagtattccttgttt
gcagtgattaatcaccatggaactttggaaagtggccactataccagcttcatccggcaacaaaaggaccagtggttcagc
tgtgatgatgccatcatcaccaaggctaccattgaggacttactctacagtgaagggtatttactgttctatcacaaacaggg
tctagagaaa (SEQ ID NO: 6)
ctgChr_Xctg264 MOOSE13941 1564654..1565727
RKKSVYTVGLRGLINLGNTCFMNCIVQALTHIPLLKDFFLSDKHKCIMTSPSL
CLVCEMSSLFHAMYSGSRTPHIPYKLLHLIWIHAEHLAGYRQQDAHEFLIAIL
DVLHRHSKDDSGGQEANNPNCCNCIIDQIFTGGLQSDVTCQACHSVSTTIDP
CWDISLDLPGSCATFDSQNPERADSTVSRDDHIPGIPSLTDCLQWFTRPEHLG
SSAKIKCNSCQSYQESTKQLTMKKLPIVACFHLKRFEHVGKQRRKINTFISFP
LELDMTPFLASTKESRMKEGQPPTDCVPNENKYSLFAVINHHGTLESGHYTS
FIRQQKDQWFSCDDAIITKATIEDLLYSEGYLLFYHKQGLEK (SEQ ID NO: 54)
agaaaaaagtcagtctatactgtaggcctgagagggctaatcaatcttgggaacacttgttttatgaattgtattgt
ccaggcacttacccatattcctctactgaaagatttcttcctctctgacaagcacaaatgtataatgacaagccccagcttgtg
tctggtctgtgaaatgtcttcgctttttcatgctatgtactctgggagccgaactcctcacattccctataagttactgcatctgat
atggatccatgcagaacatttagcagggtacaggcagcaggatgcccatgagttccttattgcaatattagacgtgctacat
agacacagcaaagatgatagtggtgggcaggaggccaataaccccaactgctgtaactgcatcatagaccaaatctttac
aggtggcctgcaatcagatgtcacatgtcaagcctgccatagtgtttctaccaccatagacccatgctgggacatcagtttg

gacttgcctggctcttgtgccacattcgattcccagaacccagagagggctgacagcacagtgagcagggatgaccacat
accaggaatcccctcacttacagactgtctacagtggtttacaaggccagagcacctaggaagcagtgccaaaatcaaat
gcaatagttgccaaagctaccaggagtctactaaacagctcacaatgaaaaaattacccattgtggcttgttttcatctcaag
cggtttgagcatgtaggcaaacagaggcgaaagattaataccttatctcctttcccttggagctggacatgactccgtttttg
gcctctactaaagagagcagaatgaaagaaggccagccaccaacagattgtgtgcccaatgagaataagtattccttgttt
gcagtgattaatcaccatggaactttggaaagtggccactataccagcttcatccggcaacaaaaggaccagtggttcagc
tgtgatgatgccatcatcaccaaggctaccattgaggacttactctacagtgaagggtatttactgttctatcacaaacaggg
tctagagaaa (SEQ ID NO: 7)

ctgChr_Xctg26 MOOSE13943 4457988..4459058

RITSSFTIGLRGLINLGNTCFMNCIVQALTHTPILRDFFLSDRHRCEMPSPELCL
VCEMSSLFRELYSGNPSPHVPYKLLHLVWIHARHLAGYRQQDAHEFLIAALD
VLHRHCKGDDVGKAANNPNHCNCIIDQIFTGGLQSDVTCQACHGVSTTIDPC
WDISLDLPGSCTSFWPMSPGRESSVNGESHIPGITTLTDCLRRFTRPEHLGSSA
KIKCGSCQSYQESTKQLTMNKLPVVACFHFKRFEHSAKQRRKITTYISFPLEL
DMTPFMASSKESRMNGQLQLPTNSGNNENKYSLFAVVNHQGTLESGHYTSF
IRHHKDQWFKCDDAVITKASIKDVLDSEGYLLFYHKQVLE (SEQ ID NO: 55)

agaatcacctccagctttacgatcggtttaagaggactcatcaatcttggcaacacgtgctttatgaactgcattgt
ccaggccctcacccacacgccgatactgagagatttctttctctctgacaggcaccgatgtgagatgccgagtcccgagtt
gtgtctggtctgtgagatgtcgtcgctgtttcgggagttgtattctggaaacccgtctcctcatgtgccctataagttactgcac
ctggtgtggatacatgcccgccatttagcagggtacaggcaacaggatgcccacgagttcctcattgcagcgttagatgtc
ctgcacaggcactgcaaaggtgatgatgtcgggaaggcggccaacaatcccaaccactgtaactgcatcatagaccaaa
tcttcacaggtggcctgcagtctgatgtcacctgtcaagcctgccatggcgtctccaccacgatagacccatgctgggaca
ttagtttggacttgcctggctcttgcacctccttctggcccatgagcccagggagggagagcagtgtgaacggggaaagc
cacataccaggaatcaccaccctcacggactgcttgcggaggtttacgaggccagagcacttaggaagcagtgccaaaa
tcaaatgtggtagttgccaaagctaccaggaatctaccaaacagctcacaatgaataaattacctgtcgttgcctgttttcattt
caaacggtttgaacattcagcgaaacagaggcgcaagatcactacatacatttcctttcctctggagctggatatgacgccg
tttatggcctcaagtaaagagagcagaatgaatggacaattgcagctgccaaccaatagtggaaacaacgaaaataagtat
tccttgtttgctgtggttaatcaccaaggaaccttggagagtggccactataccagcttcatccggcaccacaaggaccagt
ggttcaagtgtgatgatgccgtcatcactaaggccagtattaaggacgtactggacagtgaagggtatttactgttctatcac
aaacaggtgctagaa (SEQ ID NO: 8)

ctg8ctg6 MOOSE13948 967504..968427

LSSRRPAAVGAGLQNMGNTCYENASLQCLTYTPPLANYMLSREHSQTCQRP
KCCMLCTMQAHITWALHSPGHVIQPSQALAAGFHRGKQEDAHEFLMFTVD
AMKKACLPGHKQVDHHSKDTTLIHQIFGGCWRSQIKCLHCHGISDTFDPYLD
IALDIQAAQSVKQALEQLVKPEELNGENAYHCGLCLQRAPASKTLTLHTSAK
VLILVLKRFSDVTGNKLAKNVQYPECLDMQPYMSQQNTGPLVYVLYAVLV
HAGWSCHDGHYFSYVKAQEGQWYKMDDAKVTACSITSVLSQQAYVLFYIQ
KSE (SEQ ID NO: 56)

ctgagtagcaggagacctgctgcggtgggggctgggctccagaatatgggaaatacctgctacgagaacgcttccctgc
agtgcctgacatacacaccgccccttgccaactacatgctgtcccgggagcactctcaaacatgtcagcgtcccaagtgct
gcatgctctgtactatgcaagctcacatcacatgggccctccacagtcctggtcatgtcatccagccctcacaggcattggc
tgctggcttccatagaggcaagcaggaagatgcccatgaatttctcatgttcactgtggatgccatgaaaaaggcatgcctt
cccggccacaagcaggtagatcatcactctaaggacaccaccctcatccaccaaatatttggaggctgctggagatctca
aatcaagtgtctccactgccacgggatttcagacactttgacccttacctggacatcgccctggatatccaggcagctcag
agtgtcaagcaagctttggaacagttggtgaagcccgaagaactcaatggagagaatgcctatcattgcggtctttgtctcc
agagggcgccagcctccaagacgttaactttacacacttctgccaaggtcctcatccttgtcttgaagagattctccgatgtc

acaggcaacaaacttgccaagaatgtgcaatatcctgagtgccttgacatgcagccatacatgtctcagcagaacacagg
acctcttgtctatgtcctctatgctgtgctggtccacgctgggtggagttgtcacgatggacattacttctcttatgtcaaagct
caagaaggccagtggtataaaatggatgatgccaaggtcactgcctgtagcatcacttctgtcctgagtcaacaggcctat
gtcctcttttacatccagaagagtgaa (SEQ ID NO: 9)

ctgChr_1ctg34 MOOSE13952 28934534..28934662, 28935089..28935291, 28936850..28937023, 28937998..28938168, 28940593..28940826, 28943293..28943311, 28943898..28943978
CEKRENLLPFVGLNNLGNTCYLNSILQVLYFCPGFMYCIFKTRIDEMEIFIYRE
LNPMYEGYLQHDAQEVLQCILGNIQETCQLLKKEEVKNVAELPTKVEEIPHP
KEEMNGEEQIGFELVEKLFQGQLVLRTRCLECESLTERREDFQDISVPVQEDE
LSKVEESSEKMKTLRWAISQFASVERIVGEDKYFCENCHHYTEAERSLLFDK
MPEVITIHLKCFAASGNLFFFSSQRFDCYGGGLSKINTPLLTPLKLSLEEWSTK
PTNDSYGLFAVVMHSGITISSGHYTASVKEYEGKWLLFDDSEVKVTEEKDFL
NSLSPSTSPTSTPYLLFYKK (SEQ ID NO: 57)

tgtgagaagagagaaaacttgttaccatttgtgggactgaataatctcggcaatacttgctatcttaatagtatacttcaggtat
tatatttttgtcccGGCtttatgtactgtatcttcaaaactagaatagatgaaatggaaatttttatttatagggaactcaaccct
atgtatgaaggatatctacagcatgatgcacaggaagtattacaatgtattttgggaaacattcaagaaacatgccaactcct
aaaaaaagaagaagtaaaaaatgtggcagaattacctactaaggtagaagaaatacctcatccgaaagaggaaatgaat
GGTgaagaacaaattggttttgagctagtggagaaattatttcaaggtcagctggtattaaggacgcgttgcttggaatgt
gaaagtttaacagaaagaagagaagattttcaagacatcagtgtgccagtacaagaagatgagctttccaaagtagagga
gagttctgaaAAAatgaagaccctgagatgggcaatttcacaatttgcttcagtagaaaggattgtaggagaagataaat
atttctgtgaaaactgccatcattatactgaagctgaacgaagtcttttgtttgacaaaatgcctgaagttataactattcatttg
aagtgctttgctgctagtGGAaacctttttctttttttcctcccaaaggtttgattgttatggtggtggactttccaagatcaaca
ctcctttattgacacctcttaaattgtcactagaagaatggagcacaaagccaactaacgacagctatggattatttgcggttg
tgatgcatagtggcattacaattagtagtgggcattacactgcttctgttaaagagtatgaggggaagtggttgctttttgatg
attctgaagtcaaagttactgaagagaaggactttctgaattctctttccccttctacatctcctacttctactccttacttgctatt
ttataagaaa (SEQ ID NO: 10)

ctg8ctg4 MOOSE13954 6009332..6010255
LSSRRPAAVGAGLQNMGNTCYENASLQCLTYTLPLANYMLSREHSQTCQRP
KCCMLCTMQAHITWALHSPGHVIQPSQALAAGFHRGKQEDVHEFLMFTVD
AMKKACLPGHKQVDHHCKDTTLIHQIFGGCWRSQIKCLHCHGISDTFDPYL
DIALDIQAAQSVKQALEQLVKPEELNGENAYHCGLCLQRAPASNTLTLHTSA
KVLILVLKRFSDVAGNKLAKNVQYPECLDMQPYMSQQNTGPLVYVLYAVL
VHAGWSCHDGHYFSYVKAQEVQWYKMDDAEVTVCSIISVLSQQAYVLFYI
QKSE (SEQ ID NO: 58)

ctgagtagcaggagacctgctgcggtgggggctgggctccagaatatgggaaatacctgctacgagaacgcttccctgc
agtgcctgacatacacactgccccttgccaactacatgctgtcccgggagcactctcaaacatgtcagcgtcccaagtgct
gcatgctctgtactatgcaagctcacatcacatgggccctccacagtcctggccatgtcatccagccctcacaggcattgg
ctgctggcttccatagaggcaagcaggaagatgtccatgaatttctcatgttcactgtggatgccatgaaaaaggcatgcct
tcccggccacaagcaggtagatcatcactgcaaggacaccaccctcatccaccaaatatttggaggctgctggagatctc
aaatcaagtgtctccactgccacgggatttcagacacttttgacccttacctggacatcgccctggatatccaggcagctca
gagtgtcaagcaagctttggaacagttggtgaagcccgaagaactcaatggagagaatgcctatcattgcggtctttgtctc
cagagggcgccggcctccaacacgttaactttacacacttctgccaaggtcctcatccttgtcttgaagagattctccgatgt
cgcaggcaacaaacttgccaagaatgtgcaatatcctgagtgccttgacatgcagccatacatgtctcagcagaacacag
gacctcttgtctatgtcctctatgctgtgctggtccacgctgggtggagttgtcacgacggacattacttctcctatgtcaaag
ctcaagaagtccagtggtataaaatggatgatgccgaggtcactgtctgtagcatcatttctgtcctgagtcaacaggcctat

gtcctcttttacatccagaagagtgaa (SEQ ID NO: 11)
ctg21fin2 MOOSE13975 2828213..2828293, 2836386..2836472,
2845420..2845514, 2846133..2846213, 2860640..2860778, 2868030..2868242
YDRKRQDKAPVGLKNVGNTCWFSAVIQSLFNLLEFRRLVLNYKPPSNAQDL
PRNQKAFFFSQQDVSEFTHKLLDWLEDAFQMKAEEETVGKDVEKLKPLCSV
GEDMKWYSHCGKHFCYCFISFQHWFTELPPVLTFELSRFEFNQALGRPEKIH
NKLEFPQVPYRLHAVLVHEGQANAGHYWAYIFDHRESRWMKYNDIAVTKS
SWEELVRDSFGGYRNASAYCLMYINDKA (SEQ ID NO: 59)
tatgatagaaaaagacaggacaaagctcccgttgggctaaagaatgttggcaatacttgttggtttagtgctgttattcagtc
attatttaatcttttggaatttagaagattagttctgaattacaagcctccatcaaatgctcaagatttaccccgaaaccaaaag
gctttttttcttttcacagcaagatgtgagtgagtttacacacaaattattagattggttagaagatgccttccaaatgaaagctg
aagaggagACGgttggcaaggatgtggagaaattgaaacccttgtgcagtgttggtgaggatatgaaatggtacagcc
actgtggaaaaCATttttgttattgttttatttcctttcagcattggtttactgaattaccacctgtgttaacatttgaattgtcaa
gatttgaatttaatcaggcattgggaagaccagaaaaaattcacaacaaattagaatttccccaagttccttatcgattacatg
ccgttttagttcacgaaggccaagctaatgctgggcactactgggcatatatttttgatcatcgtgaaagcagatggatgaag
tacaatgatattgctgtgacaaaatcatcatgggaagagctagtgagggactcttttggtggttatagaaatgccagtgcata
ctgtttaatgtacataaatgataaggca (SEQ ID NO: 12)
ctgchr11q_5 MOOSE13977 18732295..18732375, 18738638..18738724,
18739545..18739553, 18740980..18741048, 18743645..18743770,
18745607..18745668, 18792900..18792942, 18807171..18807377
NDWRRVDGWPVGLKNVGNTCWFSAVIQSLFQLPEFRRLVLSYSLPQNVLEN
CRSHTEQQQDVSEFTHKLLDWLEDAFQLAVNTFGQYPLQVNGYRNLDECLE
GAMVEGDVELLPSDHSVKYGQERWFTKLPPVLTFELSRFEFNHSWGRDKKD
SKALHTVPYRLHAVLVHEGQANAGHYWAYIYNQPRQSWLKYNDISVTESS
WEEVERDSYGGLRNVSAYCLMYIND (SEQ ID NO: 60)
aatgactggaggagagttgatggttggccagttgggctgaaaaatgttggcaatacatgttggtttagtgctgttattcagtct
ctctttcaattgcctgaatttcgaagacttgttctcagttatagtctgccacaaaatgtacttgaaaattgtcgaagtcatacaga
acagcagcaagatgtgagtgaattcacacacaagctcctggattggctagaggacgcattccagctagctgttaataccttc
ggccagtatcctcttcaggtaaacggttatcgcaacttagacgagtgtttggaaggggccatggtggagggtgatgttgag
cttcttccctccgatcactcggtgaagtatggacaagagcgttggtttacaaagctacctccagtgttgacctttgaactctca
agatttgagtttaatCACtcctggggcagagacaagaaggattcaaaggccttgcacacagtgccttatcgcttgcatgc
agttcttgttcatgaaggacaagcaaatgctggacactattgggcctatatctataatcaaccccgacagagctggctcaag
tacaatgacatctctgttactgaatcttcctgggaagaagttgaaagagattcctatggaggcctgagaaatgttagtgcttac
tgtctgatgtacattaatgac (SEQ ID NO: 13)
ctg17005 MOOSE13980 1430240..1430370, 1438626..1438974,
1439081..1439207, 1445603..1445624, 1451804..1451827, 1452650..1452714,
1473305..1473354 SGSSPSSSWPSGLRSSCPIFQCLFMLHLLSRSQSFLWPRV
RMRRQHGALEFHRVLFGSLQEERAQDADSVWQQQQAHQQHSCTLDECFQF
YTKEEQLAQDDAWKCPHCQVLQQGMVKLSLWTLPDILIIHLKRFCQVGERR
NKLSTLVKFPLSGLNMAPHVAQRSTSPEAGLGPWPSWKQPDCLPTSYPLDFL
YDLYAVCNHHGNLQGGHYTAYCRNSLDGQWYSYDDSTVEPLREDEVNTR
GAYILFYQKRNSIP (SEQ ID NO: 61)
tctggaagctctccttccagttcatggccttctggcctccggtcgagcTGTcccatttttcaatgtttgtttatgttgcatttgc
tttcgaggtctcagtcattcctctggccaAGAgtgaggatgaggaggcagcatGGTgctctggaatttcatagaGT
Cctgttcgggagcctccaggaggagcgagcgcaggatgccgacagtgtgtggcagcagcaggcgcatcagcag
cacagctgtaccttggatgaatgttttcagttctacaccaaggaggagcagctggcccaggatgacgcctggaagtgtcct

cactgccaagtcctgcagcaggggatggtgaagctgagtttgtggacgctgcctgacatcctcatcatccacctcaaaag
gttctgccaggtgggcgagagaagaaacaagctctccacgctggtgaagtttccgctctctggactcaacatggctcccc
atgtggcccagagaagcaccagccctgaggcaggactgggcccctggccttcctggaagcagccggactgcctgccca
ccagttacccgctggacttcctgtacgacctgtatgccgtctgcaaccaccatggcaacctgcaaggtgggcattacaca
GCCtactgccggaactctctggatggccagtggtacagttatgatgacagcacggtggaaccgcttcgagaagatgag
gtcaacaccagaggggcttatatcctgttctatcagaagcggaacagcatccct (SEQ ID NO: 14)
ctgCHR12_11 MOOSE13982 65213..65322, 76770..76857, 78921..78986,
80779..80841, 88642..88701, 89241..89327, 90016..90233, 92975..93203
LPPAFFLGLVPGLVNLGNTCFMNSLLQGLSACPAFISLALFIFESLLPLYSCSFI
AQEGIHLYRQQDAHELFHVITSSLEDERDRQPRSQHPFHGRLTSNMVCKHCE
HQSPVRFDTFDSLSLSIPAATWGHPLTLDHCLHHFISSESVRDVVCDNCTKRT
TFVKQLKLGKVSPHYTPCWLCFEDSVYPAPETTRFSRFLFHPQLPQCLCIHLQ
RLSWSSHGTPLKRHEHVHSSTYLFRLMAVVVHHGDMHSGHFVTYRRSPPSA
RNPLSTSNQWLWVSDDTVRKASLQEVLSSSAYLLFYERVLSRM (SEQ ID
NO: 62)
cttccccctgctttttttttagggcttgtgcctggccttgttaatttagggaacacctgcttcatgaactccctgctacaaggcct
gtctgcctgtcctgctttcatcAGCcttgccttatttatctttgaatctctacttccactctacagctgcagttttatagcccagg
agggaattcatttatataggcaacaagatgctcacgaattattccatgtcattacctcgtcattggaagatgagcgagaccgc
cagcctcggtctcaacatccttttcatggaagactcactagtaatatggtctgcaaacactgtgaacaccagagtcctgttcg
atttgataccttgatagcctttcactaagtattccagccgccacatggggtcacccattgaccctggaccactgccttcacca
cttcatctcatcagaatcagtgcgggatgttgtgtgtgacaactgtacaaagaggaccacttttgttaaacagttaaaactag
ggaaggtgagcccacactacacaccctgttggctttgttttgaggactccgtgtatcctgcccctgaaacaactcggttctcc
cgatttctcttccacccgcagctccctcagtgtctctgcatccacctacagcggctgagctggtccagccacggcacgcct
ctgaagcggcatgagcacgtgCACtcctccacatacctcttccggctgatggcagttgtcgtccaccatggagacatgc
actctggacactttgtcacttaccgacggtccccaccttctgccaggaaccctctctcaactagcaatcagtggctgtgggt
ctccgatgacactgtccgcaaggccagcctgcaggaggtcctgtcctccagcgcctacctgctgttctacgagcgcgtcct
ttccaggatg (SEQ ID NO: 15)
ctg2013 MOOSE14044 1981177..1981331, 1981824..1981954,
1982040..1982121, 1984518..1984699, 1985403..1985520, 1985943..1985983,
2073983..2074221 MLSSPDFYPAYPSAMQSLSLGSLARALELMTQYFNNW
NWVYDNIIDQNESKLSKSRREEIERDRERKERREGDREKKRQNGVVEVPFLL
SSKYDEPSRQVILEALAEFERSTCIRFVTYQDQRDFISIIPMYGCFSSVGRSGG
MQVVSLAPTCLQKGRGIVLHELMHVLGFWHEHTRADRDRYIRVNWNEILP
GFEINFIKSRSSNMLTPYDYSSVMHYGRLAFSRRGLPTITPLWAPSVHIGQRW
NLSASDITRVLQLYGCSPKLEKVNTVNIKIIFLYTQLENTIVSKHTIIIAIIKVPR
NKSDSCIKYLRKNE (SEQ ID NO: 63)
atgctctcctcccccgatttctacccagcttacccctctgcaatgcaatctctaagccttggctccttggccagggccttagaa
ctgatgacccagtactttaacaattggaactgggtctacgacaacataatagatcagaatgaaagcaaattgagtaagtcaa
ggagagaagagatagagagagacagagagagaaaagagagaagagagggagacagggaaaaaaagagacagAA
TggtgtcgtggaggtccccttcctgctctccagcaagtacGATgagcccagccgccaggtcatcctggaggctcttgc
ggagtttgaacgttccacgtgcatcaggtttgtcacctatcaggaccagagagacttcatttccatcatccccatgtatGG
Gtgcttctcgagtgtggggcgcagtggagggatgcaggtggtctccctggcgcccacgtgtctccagaagggccgggg
cattgtccttcatgagctcatgcatgtgctgggcttctggcacgagcacacgcgggccgaccgggaccgctatatccgtgt
caactggaacgagatcctgccaGGCtttgaaatcaacttcatcaagtctcggagcagcaacatgctgacgccctatgac
tactcctctgtgatgcactatgggAGGcttgccttcagccggcgtgggctgcccaccatcacaccactttgggccccca
gtgtccacatcggccagcgatggaacctgagtgcctcggacatcacccgggtcctccaactctacggctgcagcccaA

AAttagaaaaggtaaatacagttaatataaaaatcatatttctgtacacccagttagaaaacacaattgttagtaaacatacc
attataatagcaatcataaaggtcccaagaaataaatctgacagctgtatcaaatatttgaggaaaaatgaa (SEQ ID NO: 16)

ctgCHR12_3 MOOSE14065 6845696..6845761, 6845907..6846079, 6849800..6849892, 6852604..6852721, 6858283..6858419, 6858523..6858704, 6859916..6860061, 6868904..6869011, 6870600..6870741, 6872661..6872804, 6874609..6874714, 6896403..6896443, 6898513..6898637, 6899178..6899261, 6908160..6908289, 6919052..6919123, 6920651..6920743, 6925041..6925098, 6926345..6926386, 6929101..6929215

FQLWIWLRPCPVTWIPRFPGGGVFPGGSLSPLHILGTKAFFKVLFLDHHFRLY
MEHSNDIISPHFKEITQIITSFQEIIEEEFGISQCYTYNNPSKSDIRIHWTVSDLSQ
VFILSRAKSLYIQILSQRHSRKKRLISYPRYIEIMVTADAKVVSAHGSNLQNYI
LTLMSIVATIYKDPSIGNLIHIVVVKLVMIHREEEGPVINFDGATTLKNFCSWQ
QTQNDLDDVHPSHHDTAVLITREDICSSKEKCNMLGLSYLGTICDPLQSCFIN
EEKGLISAFTIAHELGHTLGVQHDDNPRCKEMKVTKYHVMAPALSFHMSPW
SWSNCSRKYVTEFLDTGYGECLLDKPDEEIYNLPSELPGSRYDGNKQCELAF
GPGSQMCPHIENICMHLWCTSTEKLHKGCFTQHVPPADGTDCGPGMHCRH
GLCVNKETETRPVNGEWGPWEPYSSCSRTCGGGIESATRRCNRPEPRNGGN
YCVGRRMKFRSCNTDSCPKGTQDFREKQCSDFNGKHLDISGIPSNVRWLPR
YSGIGTKDRCKLYCQVAGTNYFYLLKDMVEDGTPCGTETHDICVQGQCMA
AGCDHVLNSSAKIDKCGVCGGDNSSCKTITGVFNSSHYGYNVVVKIPAGAT
NVDIRQYSYSGQPDDSYLALSDAEGNFLFNGNFLLSTSKKEINVQGTRTVIEY
SGSNNAVERINSTNRQEKELILQVLCVGNLYNPDVHYSFNIPLEE (SEQ ID NO: 64)

ttccagctctggatatggttaaggccttgccctgtgacctggattcccaggtttcctggtgggggtgtatttcctggaggcagt
ctctccccccttgcacattttggggaccaaaGCAtttttaaagtgctcttcctggaccaccattttagactaTATatggag
catagtaatgacataatcagtccacattttaaagagatcactcagattATCacatcttttcaagaaataatagaggaagagt
ttggaatatcacaatgctatacatataataatccatctaaatctgacattaggatccatTGGacagtatcagatttgtctcag
gtttttattctaagtagagccaaaagcctttatatccaaatcttgtctCAAagacattccaggaaaaaacgtcttatatcatat
ccaagatacattgaaattatggttacagctgatgctaaagtggtttctgctcatggatcgaatttgcaaaactatatactgactc
taatgtcaattgttgcaacaatctacaaagatccaagtattggaaatttgatacacatagtagtggtaaaattagttatgattca
ccgtgaggaggaaggaccagtcattaattttgatggtgctaccacattaaagaacttttgttcatggcaacaaactcagaatg
accttgatgatgttcacccttcccaccatgacactgctgttcttatcactAGGgaagacatttgttcatctaaagagaaatgt
aacatgttaGGTttatcatatttaggtaccatatgtgatcctttacaaagctgctttattaatgaagaaaaaggactcatttctg
cttttactatagcccatgagcttgggcacACActtggtgttcaacatgatgataatcctagatgtaaagaaatgaaagttac
aaagtatcatgtaatggcccctgctttaagttttcacatgagtccttggagctggtcaaactgtagtcggaaatatgttactga
attcctaGATactggttacggggaatgtcttcttgacaaaccagatgaagaaatatataatctgccttcagaacttcctgga
tcacgatatgatggaaacaagcagtgtgagcttgcgtttggtcctgggtcacaaatgtgtccccatatagagaatatatgcat
gcatctgtggtgcacaagcacagaaaagcttcacaaaggctgtttcactcaacacgtgccaccagcagatggaacagact
gcggtcctggaatgcattgccgtcatgggctatgtgtaaacaaagaaacggaaacacgtcctgtaaatggtgaatgggga
ccatgggaaccttacagttcttgttcaagaacatgtggaggcggaatcgaaagtgcaaccaggcgctgtaatcgtcctGA
Gccaagaaacggaggaaattactgtgtgggccgcaggatgaaatttcgatcatgtaatactgattcatgtccaaaaggca
cacaagactttcgagagaagcagtgctctgattttaatggtaaacatttggacatcagtggcattccctctaatgtgaggtgg
cttccaagatacagtggcATTggcacaaaggatcgttgtaaactctattgtcaggttgctggaaccaattatttctacctatt
gaaggatatggttgaagatggtactccttgtggaactgaaactcatgacatctgtgttcaaggccagtgtatggcagctggtt
gtgatcacgtgttaaactccagtgccaagatagacaaatgtggagtgtgtggtggggacaactcttcatgcaagacaataa

caggtgtcttcaacagttctcattatGGTtataatgttgttgtaaagattcccgcaggagcaacaaacgttgacattcgtca
gtacagctattctggacaaccagatgacagttaccttGCAttatctgacgctgaagggaattttcttttcaatggaaattttct
tctaagtacgtcaaaaaaagaaatcaatgtgcaaggaacaagaactgttattgaatacagtggatcaaataacgcagttgaa
agaattaatagtactaatcgacaagagaaagaacttattttgcaggtgttgtgtgtgggtaatttatacaaccctgatgtacatt
attccttcaatatccctttggaagag
ctg10007 MOOSE14068 24799456..24799495, 24807076..24807253,
24811769..24811846, 24812886..24813009, 24813394..24813527,
24815913..24816088, 24818085..24818233, 24820282..24820395,
24821774..24821906, 24823047..24823190, 24824716..24824821,
24826817..24826973, 24827698..24827777, 24848289..24848468,
24848491..24848535, 24848850..24848955, 24849153..24849274,
24849572..24849605 PSLLSCLLSFPRPGPDIAWQLSCKGSWIGTQTHSLTVSS
ATDFVLRWQNRMVEYPGVPQMPYGGHSSPMTFLLYGDIANFDFYSNLVVT
APPVGWTSLSSCLDLPNLLGLVGDQLGDTERKRRHAKPGSYSIEVLLVVDDS
VVRFHGKEHVQNYVLTLMNIVNEIYHDESLGVHINIALVRLIMVGYRQSLSLI
ERGNPSRSLEQVCRWAHSQQRQDPSHAEHHDHVVFLTRQDFGPSGMQGYA
PVTGMCHPLRSCALNHEDGFSSAFVIAHETGHVLGMEHDGQGNGCADETSL
GSVMAPLVQAAFHRFHWSRCSKLELSRYLPSYDCLLDDPFDPAWPQPPELP
GINYSMDEQCRFDFGSGYQTCLAFRTFEPCKQLWCSHPDNPYFCKTKKGPP
LDGTECAPGKWCFKGHCIWKSPEQTYGQDGGWSSWTKFGSCSRSCGGGVR
SRSRSCNNPSPAYGGRLCLGPMFEYQVCNSEECPGTYEDFRAQQCAKRNSY
YVHQNAKHSWVPYEPDDDAQKCELICQSADTGDVVFMNQVVHDGTRCSY
RDPYSVCARGECVPVGCDKEVGSMKADDKCGVCGGDNSHCRTVKGTLGK
ASKQAALKLVQIPAGARHIQIEALEKSPHRIVVKNQVTGSFILNPKGKEATSR
TFTAMGLEWEDAVEDAKESLKTSGPLPEAIAILVSPTLDTQNIKEPRHRPD cc
cagcctcctgtcctgccttcttagctttcccCGCccagggccagatatagcctggcagctgtcttgtaaaggaagttggat
tggaacacagacacactcactcacagtctcatctgcgactgattttgtgcttcggtggcagaatcgaatggttgagtatcccg
gagttcctcagatgccctatggaggccacagtagccccatgacttttctcctttatggagatattgccaactttgatttttacag
caaccttgtgGTGacagcccctcctgtgggttggacctcactctcttcctgtcttGACcttcccaacctgctgggcctgg
tgggggaccagctgggcgacacagagcggaagcggcggcatgccaagccaggcagctacagcatcgaggtgctgct
ggtggtggacgactcggtggttcgcttccatggcaaggagcatgtgcagaactatgtcctcaccctcatgaatatcgtgA
ATgagatttaccacgatgagtccctgggggttcatataaatattgccctcgtccgcttgatcatggttggctaccgacagtc
cctgagcctgatcgagcgcgggaacccctcacgcagcctggagcaggtgtgtcgctgggcacactcccagcagcgcca
ggaccccagccacgctgagcaccatgaccacgttgtgttcctcacccggcaggactttgggccctcaggtatgcaaGG
Gtatgcacccgtcactggcatgtgtcacccccctgaggagctgtgccctcaaccatgaggatggcttctcctcagccttcgt
gatagctcatgagaccggccacGTGctcggcatggagcatgacggtcaggggaatggctgtgcagatgagaccagc
ctgggcagcgtcatggcgcccctggtgcaggctgccttccaccgcttccattggtcccgctgcagcaagctggagctcag
ccgctacctcCCCtcctacgactgcctcctcgatgacccctttgatcctgcctggccccagcccccagagctgcctggg
atcaactactcaatggatgagcagtgccgctttgactttggcagtggctaccagacctgcttggcattcaggacctttgagc
cctgcaagcagctgtggtgcagccatcctgacaacccgtacttctgcaagaccaagaaggggcccccgctggatgggac
tgagtgtgcacccggcaagtggtgcttcaaaggtcactgcatctggaagtcgccggagcagacatatggccaggatgga
ggctggagctcctggaccaagtttgggtcatgttcgcggtcatgtgggggcggggtgcgatcccgcagccggagctgca
acaaccccTCCccagcctatggaggccgcctgtgcttagggcccatgttcgagtaccaggtctgcaacagcgaggagt
gccctgggacctacgaggacttccgggcccagcagtgtgccaagcgcaactcctactatgtgcaccagaatgccaagca
cagctgggtgccctacgagcctgacgatGACgcccagaagtgtgagctgatctgccagtcggcggacacgggggac
gtggtgttcatgaaccaggtggttcacgatgggacacgctgcagctaccgggacccatacagcgtctgtgcgcgtggcg

agtgtgtgcctgtcggctgtgacaaggaggtggggtccatgaaggcggatgacaagtgtggagtctgcgggggtgacaa
ctcccactgcaggactgtgaaggggacgctgggcaaggcctccaagcaggcaGCTctcaagctggtgcagatccca
gcaggtgccaggcacatccagattgaggcactggagaagtccccccaccgcattGTGgtgaagaaccaggtcaccg
gcagcttcatcctcaaccccaagggcaaggaagccacaagccggaccttcaccgccatgggcctggagtgggaggatg
cggtggaggatgccaaggaaagcctcaagaccagcgggcccctgcctgaagccattgccatcctggtgagccccactC
TTgatacacaaaatattaaagaacccaggcacagaccagac
ctgC22 MOOSE14071 8022899..8022973, 8038812..8039047,
8039484..8039515, 8040059..8040156, 8041290..8041420, 8044820..8044992,
8054710..8054858, 8060658..8060753, 8072693..8072840, 8074894..8075031,
8078272..8078377, 8081059..8081218, 8082756..8082839, 8083142..8083427,
8084423..8084440, 8085600..8085614, 8087906..8088011, 8105355..8105431,
8108468..8108544, 8118599..8118640
LLLLLLLLLFLRQSWQGPIISATQEAEAVESLEPRRRRLHSGVRDQPGQHGET
LSLLKIQKLDRHGLIRTRKNEFLISPLPQLLAQEHNYSSPAGHHPHVLLDCFY
HCHIKDFSSSLVSVSLSTVLSRYSLISLPNALLFIVDAPKPPTEDTYLRFDEYGS
SGRPRRSAGKSQKGLNVETLVVADKKMVEKHGKGNVTTYILTVMNMVSGL
FKDGTIGSDINVVVVSLILLEQEPGGLLINHHADQSLNSFCQWQSALIGKNGK
RHDHAILLTGFDICSWKNEPCDTLGFAPISGMCSKYRSCTINEDTGLGLAFTI
AHESGHNFGMIHDGEGNPCRKAEGNIMSPTLTGNNGVFSWSSCSRQYLKKF
LSTPQAGCLVDEPKQAGQYKYPDKLPGQIYDADTQCKWQFGAKAKLCSLG
FDICKSLWCHRVGHRCETKFMPAAEGTVCGLSMWCRQGQCVKFGELGPRPI
HGQWSAWSKWSECSRTCGGGVKFQERHCNNPKPQYGGLFCPGSSRIYQLC
NINPCNENSLDFRAQQCAEYNSKPFRGWFYQWKPYTKVEEEDRCKLYCKA
ENFEFFFAMSGKVKDGTPCSPNKNDVCIDGVCELVGCDHELGSKAVSDACG
VCKGDNSTCKFYKGLFKQFSCLTLLKYYPVVLIPAGARSIEIQELQVSSSYLA
VRSLSQKYYLTGGWSIDWPGEFPFAGTTFEYQRSFNRPERLYAPGPTNETLC
CSVAQAGGQLRDLGSLQAPPPEFT ttgttgttgttgttgttgttgttgtttttgagacagagtt
ggcagggacctataatctcagcgactcaggaggctgaggcagtagaatcacttgaacccaggaggcggaggttgCAT
tcaggagttcgagaccagcctggtcaacatggtgaaaccctgtctctactgaaaatccaaaaattagacaggcatGGTtt
aataaggacacgaaaaaatgaattcctcatctcgccattacctcagcttctggcccaggaacacaactacagctcccctgc
gggtcaccatcctcacgtaCTAttggattgctttTATcattgtcacatcaaaGATttctcttcttcccttgtttcagtttct
ctctcaactgtcttatccagatattctctaatatcccttccaaatgctcttctgttcatcgtagatgctcccaagcctcccacaga
ggacacctatctaaggtttgatgaatatgggagctctgggcgacccagaagatcagctggaaaatcacaaaagggcctca
atgtggaaaccctcgtggtggcagacaagaaaatggtggaaaagcatggcaagggaaatgtcaccacatacattctcac
agtaatgaacatggtttctggcctatttaaagatgggactattggaagtgacataaacgtggttgtggtgagcctaattcttct
ggaacaagaacctggaggattattgatcaaccatcatgcagaccagtctctgaatagtttttgtcaatggcagtctgccctca
ttggaaagaatggcaagagacatgatcatgccatcttactaacaggatttgatatttgttcttggaagaatgaaccatgtgac
actctaGGGtttgccccccatcagtggaatgtgctctaagtaccgaagttgtaccatcaatgaggacacaggacttggcctt
gccttcaccatcgctcatgagtcagggcacAACtttggtatgattcacgatggagaagggaatccctgcagaaaggctg
aaggcaatatcatgtctcccacactgaccggaaacaatggagtgttttcatggtcttcctgcagccgccagtatctcaagaa
attcctcAGCacacctcaggcggggtgtctagtggatgagcccaagcaagcaggacagtataaatatccggacaaact
accaggacagatttatgatgctgacacacagtgtaaatggcaatttggagcaaaagccaagttatgcagccttggttttgata
tttgcaaatcactttggtgccaccgagtaggccacaggtgtgagaccaagtttatgcccgcagcagaagggaccgtttgtg
gcttgagtatgtggtgtcggcaaggccagtgcgtaaagtttggggagctcgggccccggcccatccacggccagtggtc
cgcctggtcgaagtggtcagaatgttcccggacatgtggtggaggagtcaagttccaggagagacactgcaataacccc
AAGcctcagtatggtggcttattctgtccaggttctagccgtatttatcagctgtgcaatattaacccttgcaatgaaaatag

cttggattttcgggctcaacagtgtgcagaatataacagcaaaccttccgtggatggttctaccagtggaaaccctatacaa
aagtggaaGAGgaagatcgatgcaaactgtactgcaaggctgagaactttgaatttttttttgcaatgtccggcaaagtga
aagatggaactccctgctccccaaacaaaaatgatgtttgtattgacggggtttgtgaactagtgggatgtgatcatgaacta
ggctctaaagcagtttcagatgcttgtggcgtttgcaaaggtgataattcaacttgcaagttttataaaggcCTGttcaagc
aattctcctgcctcaccctcttgAAAtattatccggtggtcctcattccagctggcgcccgaagcatcgaaatccaggag
ctgcaggtttcctccagttacctcgcagttcgaagcctcagtcaaaagtattacctcaccgggggctggagcatcgactgg
cctggggagttccccttcgctgggaccacgtttgaataccagcgctctttcaaccgcccggaacgtctgtacgcgccagg
gcccacaaatgagacgctgtgttgctctgtcgcccaggctggagggcagttgcgtgatctcggctcactgcaagctccgc
ctcctgagttcacg

ctg5007 MOOSE14072 990670..990921, 996695..997045, 1032626..1032735,
1036728..1036833, 1038307..1038390, 1040553..1040712, 1042267..1042372,
1050714..1050851, 1059675..1059832, 1073384..1073466, 1082939..1083087,
1085587..1085759, 1087530..1087660, 1089712..1089876
LLLWRCPLSPAFPLLPSRLRLSAPLTSPPPLCSLSLHSPLLGPLTPPPSPPPLLSP
LPAPRSPTAPAAPAAAAATPPPAPHGASPLLTLLISEYDLVSAYEVDHRGDY
VSHEIMHHQRRRRAVAVSEVESLHLRLKGPRHDFHMDLRTSSSLVAPGFIVQ
TLGKTGTKSVQTLPPEDFCFYQGSLRSHRNSSVALSTCQGLSGMIRTEEADY
FLRPLPSHLSWKLGRAAQGSSPSHVLNEELNVETLVVVDKKMMQNHGHENI
TTYVLTILNMVSALFKDGTIGGNINIAIVGLILLEDEQPGLVISHHADHTLSSFC
QWQSGLMGKDGTRHDHAILLTGLDICSWKNEPCDTLGFAPISGMCSKYRSC
TINEDTGLGLAFTIAHESGHNFGMIHDGEGNMCKKSEGNIMSPTLAGRNGVF
SWSPCSRQYLHKFLSTAQAICLADQPKPVKEYKYPEKLPGELYDANTQCKW
QFGEKAKLCMLDFKKATLWCHRIGRKCETKFMPAAEGTICGHDMWCRGG
QCVKYGDEGPKPTHGHWSDWSSWSPCSRTCGGGVSHRSRLCTNPKPSHGG
KFCEGSTRTLKLCNSQKCPRDSVDFRAAQCAEHNSRRFRGRHYKWKPYTQ
VEDQDLCKLYCIAEGFDFFFSLSNKVKDGTPCSEDSRNVCIDGICERVGCDN
VLGSDAVEDVCGVCNGNNSACTIHRGLYTKHHHTNREYFRAACKPWAKK

ctcctcctctggcgctgcccgctgtctcccgccttccctctgctcccctctcgcctccgcctcagcgccccgctg
acctcgcctcctcccctctgctctttgtccctgcactctcccctcctcggtcctctgaccccccgccctcacctcctccctc
ctctctcccctgcccgccccgcgctctcccaccgctcccgccgccccgccgccgcggctgccactccgccccccgcg
ccgcacggagcttcacctctcctcactcttttgatttcagaatatgacctggtctctgcctacgaggttgaccacaggggcga
ttacgtgtcccatgaaatcatgcaccatcagcggcggagaagagcagtggccgtgtccgaggttgagtctcttcaccttcg
gctgaaaggccccaggcacgacttccacatggatctgaggacttccagcagcctagtggctcctggctttattgtgcagac
gttgggaaagacaggcactaagtctgtgcagactttaccgccagaggacttctgtttctatcaaggctctttgcgatcacaca
gaaactcctcagtggccctttcaacctgccaaggcttgtcaggcatgatacgaacagaagaggcagattacttcctaaggc
cacttccttcacacctctcatggaaactcggcagagctgcccaaggcagctcgccatcccacgtaCTAaatgaagaact
gaacgtggagaccttggtggtggtcgacaaaaagatgatgcaaaaccatggccatgaaaatatcaccacctacgtgctca
cgatactcaacatggtatctgctttattcaaagatggaacaataggaggaaacatcaacattgcaattgtaggtctgattcttc
tagaagatgaacagccaggactggtgataagtcaccacgcagaccacaccttaagtagcttctgccagtggcagtctgga
ttgatggggaaagatgggactcgtcatgaccacgccatcttactgactggtctggatatatgttcctggaagaatgagccct
gtgacactttgGGAtttgcacccataagtggaatgtgtagtaaatatcgcagctgcacgattaatgaagatacaggtcttg
gactggccttcaccattgcccatgagtctggacacAACtttggcatgattcatgatggagaagggaacatgtgcaaaaa
gtccgagggcaacatcatgtcccctacattggcaggacgcaatggagtcttctcctggtcaccctgcagccgccagtatct
acacaaatttctaAGCaccgctcaagctatctgccttgctgatcagccaaagcctgtgaaggaatacaagtatcctgaga
aattgccaggagaattatatgatgcaaacacacagtgcaagtggcagttcggagagaaagccaagctctgcatgctggac
tttaaaaaggcaACCctgtggtgccatcgtattggaaggaaatgtgagactaaatttatgccagcagcagaaggcacaa

tttgtgggcatgacatgtggtgccggggaggacagtgtgtgaaatatggtgatgaaggccccaagcccacccatggcca
ctggtcggactggtcttcttggtccccatgctccaggacctgcggaggggggagtatctcataggagtcgcctctgcaccaa
ccccAAGccatcgcatggagggaagttctgtgagggctccactcgcactctgaagctctgcaacagtcagaaatgtcc
ccgggacagtgttgacttccgtgctgctcagtgtgccgagcacaacagcagacgattcagagggcggcactacaagtgg
aagccttacactcaagtagaaGATcaggacttatgcaaactctactgtatcgcagaaggatttgatttcttctttttctttgtca
aataaagtcaaagatgggactccatgctcggaggatagccgtaatgtttgtatagatgggatatgtgagagagttggatgtg
acaatgtccttggatctgatgctgttgaagacgtctgtggggtgtgtaacgggaataactcagcctgcacgattcacaggg
gtctctacaccaagcaccaccacaccaaccgtgagtactttagagctgcctgcaagccttgggcaaagaag
ctgChr_1ctg2 MOOSE14097 9792..9938, 17966..18005, 19452..19616,
19867..19947, 22217..22312, 22920..23038, 23663..23728, 25005..25139,
25313..25402, 30489..30625, 30979..31072, 32590..32718, 33172..33270,
33920..34003, 34386..34489, 34836..34894, 35346..35465, 36313..36446,
36559..36624, 37252..37347, 37646..37716, 38190..38286
MSSYRPRVISRMPHLLRLLLAVTVSQTFIVFLFVFLFFVILTVLRITDQVSEVC
TTPGCVIAAAARILQNMDPTTEPCDDFYQFACGGWLRRHVIPETNSRYSIFD
VLRDELEVILKAVLENSTAKDRPAVEKARTLYRSCMNQSVIEKRGSQPLLDI
LEVVGGWPVAMDRWNETVGLEWELERQLALMNSQFNRRVLIDLFIWNDD
QNSSRHIIYIDQPTLGMPSREYYFNGGSNRKVREAYLQFMVSVATLLREDAN
LPRDSCLVQEDMVQVLELETQLAKATVPQEERHDVIALYHRMGLEELQSQF
GLKGFNWTLFIQTVLSSVKIKLLPDEEVVVYGIPYLQNLENIIDTYSARTIQNY
LVWRLVLDRIGSLSQRFKDTRVNYRKALFGTMVEEVRWRECVGYVNSNME
NAVGSLYVREAFPGDSKSMVRELIDKVRTVFVETLDELGWMDEESKKKAQ
EKAMSIREQIGHPDYILEEMNRRLDEEYSNLNFSEDLYFENSLQNLKVGAQR
SLRKLREKVDPNLWIIGAAVVNAFYSPNRNQIVFPAGILQPPFFSKEQPQALN
FGGIGMVIGHEITHGFDDNGRNFDKNGNMMDWWSNFSTQHFREQSECMIY
QYGNYSWDLADEQNVNGFNTLGENIADNGGVRQAYKAYLKWMAEGGKD
QQLPGLDLTHEQLFFINYAQVWCGSYRPEFAIQSIKTDVHSPLKVLGSLQNLA
AFADTFHCARGTPMHPKERCRVW atgtcctcttaccgtcccagggtcatatccagaatgcca
catcttctcaggctcctcttggctgtgacagtttctcagactttcattgttttcttgtttgtgtttttgttttttgtgatcttgacagtctt
gaggattactgatcaggtgagcgaggtctgcaccacccctggctgcgtgatagcaGCAgctgccaggatcctccaga
acatggacccgaccacggaaccgtgtgacgacttctaccagtttgcatgcggaggctggctgcggcgccacgtgatccct
gagaccaactcaagatacagcatctttgacgtcctccgcgacgagctggaggtcatcctcaaaGCGgtgctggagaatt
cgactgccaaggaccggccggctgtggagaaggccaggacgctgtaccgctcctgcatgaaccagAGTgtgataga
gaagcgaggctctcagcccctgctggacatcttggaggtggtgggaggctggccggtggcgatggacaggtggaacga
gaccgtaGGActcgagtgggagctggagcggcagctggcgctgatgaactcacagttcaacaggcgcgtcctcatcg
acctcttcatctggaacgacgaccagaactccagccggcacatcatctacatagaccagcccaccttgggcatgccctcc
cgagagtactacttcaacggcggcagcaaccggaaggtgcgggaagcctacctgcagttcatggtgtcagtggccacgt
tgctgcgggaggatgcaaacctgcccagggacagctgcctggtgcaggaggacatggtgcaggtgctggagctggag
acacagctggccaaggccacggtaccccaggaggagagacacgacgtcatcgccttgtaccaccggatgggactgga
ggagctgcaaagccagtttggcctgaagggatttaactggactctgttcatacaaactgtgctatcctctgtcaaaatcaagc
tgctgccagatgaggaagtggtggtctatggcatcccctacctgcagaaccttgaaaacatcatcgacacctactcagcc
AGGaccatacagaactacctggtctggcgcctggtgctggaccgcattggtagcctaagccagagattcaaggacaca
cgagtgaactaccgcaaggcgctgtttggcacaatggtggaggaggtgcgctggcgtgaatgtgtgggctacgtcaaca
gcaacatggagaacgccgtgggctccctctacgtcagggaggcgttccctggagacagcaagagcatggtcagagaac
tcattgacaaggtgcggacagtgtttgtggagacgctggacgagctgggctggatggacgaggagtccaagaagaagg
cgcaggagaaggccatgagcatccgggagcagatcgggcaccctgactacatcctggaggagatgaacaggcgcctg

gacgaggagtactccaatctgaacttctcagaggacctgtactttgagaacagtctgcagaacctcaaggtgggcgccca
gcggagcctcaggaagcttcgggaaaaggtggacccaaatCTCtggatcatcggggcggcggtggtcaatgcgttct
actccccaaaccgaaaccagattGTAttccctgccgggatcctccagccccccttcttcagcaaggagcagccacagg
ccttgaactttggaggcattgggatggtgatcgggcacgagatcacgcacggctttgacgacaatGGCcggaacttcg
acaagaatggcaacatgatggattggtggagtaacttctccacccagcacttccgggagcagtcagagtgcatgatctacc
agtacggcaactactcctgggacctggcagacgaacagaacgtgaacggattcaacacccttggggaaaacattgctga
caacggaggggtgcggcaagcctataaggcctacctcaagtggatggcagagggtggcaaggaccagcagctgcccg
gcctggatctcacccatgagcagctcttcttcatcaactatgcccaggtgtggtgcgggtcctaccggcccgagttcgccat
ccaatccatcaagacagacgtccacagtcccctgAAGgtactggggtcgctgcagaacctggccgccttcgcagaca
cgttccactgtgcccggggcacccccatgcaccccaaggagcgatgccgcgtgtgg
ctg2019 MOOSE14111 40989203..40989313, 41002051..41002170,
41003054..41003155, 41005892..41005996, 41007222..41007332,
41008443..41008533, 41010003..41010205, 41015854..41015938,
41016765..41016931, 41076010..41076027
MTALDRACLYWLFLFKLLVIDIKNNGHFYVTLANSKHLSLDFIVHITISILVKA
IQRVSRKFQTFPHFPVFYWALQTIYEWMREISEKYKEVVTQHFLGVTYETHP
MYYLKISQPSGNPKKIIWMDCGIHAREWIAPAFCQWFVKEILQNHKDNSSIR
KLLRNLDFYVLPVLNIDGYIYTWTTDRLWRKSRSPHNNGTCFGTDLNRNFN
ASWCSIGASRNCQDQTFCGTGPVSEPETKAVASFIESKKDDILCFLTMHSYG
QLILTPYGYTKNKSSNHPEMIQVGQKAANALKAKYGTNYRVGSSADILYAS
SGSSRDWARDIGIPFSYTFELRDSGTYGFVLPEAQIQPTCEETMEAVLSVLDD
LQKNPY atgactgcacttgaccgtgcttgtctttattggctgtttttatttaaattattagttattgatataaaaaataa
tggtcacttttatgtaactctcgccaactcaaaacatcttagcctggactttattgtccatatcactatcagtatttttggtcaaagc
cattcaacgagtctctaggaagttccaaactttccacattttcctgtcttctactgggcccttcaaactatctatgagtggatga
gagagatcagtgagaagtacaaggaagtggtgacacagcatttcctaggagtgacctatgagacccacccccatgtattatc
tgaagatcagccaaccatctggtaatcccaagaaaatcatttggatggactgtggaattcacgccagagaatggattgctcc
tgctttttgccaatggttcgtcaaagaaattctacaaaaccataaagacaactcaagtatacgcaagctccttaggaacctgg
acttctatgtccttccagttcttaacatagatggttatatctacacttggacaactgatcgtctttggaggaaatcccgttcaccc
cataataatggcacatgttttgggacggatctcaatcgaaatttcaatgcatcttggtgtAGTattggtgcctctagaaactg
ccaagatcaaacattctgtgggacagggccagtgtctgaaccagagactaaagctgttgccagcttcatagagagcaaga
aggatgatattttgtgcttcctgaccatgcactcttatgggcagttaattctcacaccttacggctacaccaaaaataaatcaa
gtaaccacccagaaatgattcaagttggacagaaggcagcaaatgcattgaaagcaaagtatggaaccaattatagagtt
ggatcgagtgcagatattttaTATgcctcatcagggtcttcaagagattgggcccgagacattgggattcccttctcatat
acgtttgagctgagggacagtggaacatatgggtttgttctgccagaagctcagatccagcccacctgtgaggagaccat
ggaggctgtgctgtcagtcctggatgatctacaaaaaaacccatat
ctg2019 MOOSE14139 2792089..2792135, 2796149..2796193,
2827337..2827397, 2829453..2829602, 2831346..2831494, 2832006..2832048,
2836649..2836812, 2843539..2843636, 2844957..2845004, 2845870..2845948,
2846070..2846139 PNTQNHMPLCLELGIRSYHSGFCQDCFRRNEDISHSIVL
PAAVSSAHPVPKHIKKPDYVTTGIVPDWGDSIEVKNEDQIQGLHQACQLARH
VLLLAGKSLKVDMTTEEIDALVHREIISHNAYPSPLGYGGFPKSVCTSVNNVL
CHGIPDSRPLQDGDIINIDVTVYYNGYHGDTSETFLVGNVDECGKKLVEVAR
RCRDEAIAACRAGAPFSVIGNTISHITHQNGFQVCPHFVGHGIGSYFHGHPEI
WHHANDSDLPMEEGMAFTIEPIITEGSPEFKVLEDAWTVVSLDNQRSAQFEH
TVLITSRGAQILTKLPH cctaacacacaaaaccatatgcctctgtgtctagagttaggcatcAGGagt
tatcactctgggttttgccaagattgctttagaaggaatGAAgatatttcacacagtatagttttgccggctgcagtttcttca

gctcatccggttcctaagcacataaagaagccagactatgtgacgacaggcattgtaccagactggggagacagcatag
aagttaagaatgaagatcagattcaagggcttcatcaggcttgtcagctggcccgccacgtcctcctcttggctgggaaga
gtttaaaggttgacatgacaactgaagagatagatgctcttgttcatcgggaaatcatcagtcataatgcctatccctcacctc
taggctatggaggttttccaaaatctgtttgtacctctgtaaacaacgtgctctgtcatggtattcctgacAGTcgacctctt
caggatggagatattatcaacattgatgtcacagtctattacaatggctaccatggagacacctctgaaacatttttggtggg
caatgtggacgaatgtggtaaaaagttagtggaggttgccaggaggtgtagagatgaagcaattgcagcttgcagagcag
gggctcccttctctgtaattggaaacacaatcAGCcacataactcatcagaatggttttcaagtctgtccacattttgtggg
acatggaataggatcttactttcatggacatccagaaatttggcatcatGCAaacgacagtgatctacccatggaggagg
gcatggcattcactataGAGccaatcatcacggagggatcccctgaatttaaagtcctggaggatgcatggactgtggtc
tccctagacaatcaaAGGtcggcgcagttcgagcacacggttctgatcacgtcgaggggcgcgcagatcctgaccaa
actaccccat

ctg_2 MOOSE14190 2756113..2756274, 2759580..2759716, 2761344..2761618, 2764429..2764565 RIVGGMEASPGEFPWQASLRENKEHFCGAAIINARWLV
SAAHCFNEFQDPTKWVAYVGATYLSGSEASTVRAQVVQIVKHPLYNADTA
DFDVAVLELTSPLPFGRHIQPVCLPAATHIFPPSKKCLISGWGYLKEDFLVKPE
VLQKATVELLDQALCASLYGHSLTDRMVCAGYLDGKVDSCQGDSGGPLVC
EEPSGRFFLAGIVSWGIGCAEARRPGVYARVTRLRDWILEATTKA aggatcgtg
ggcggcatggaagcatccccgggggagtttccgtggcaagccagccttcgagagaacaaggagcacttctgtggggcc
gccatcatcaacgccaggtggctggtgtctgctgctcactgcttcaatGAGttccaagacccgacgaagtgggtggcct
acgtgggtgcgacctacctcagcggctcggaggccagcaccgtgcgggcccaggtggtccagatcgtcaagcacccc
ctgtacaacgcggacacggccgactttgacgtggctgtgctggagctgaccagccctctgcctttcggccggcacatcca
gcccgtgtgcctcccggctgccacacacatcttcccacccagcaagaagtgcctgatctcaggctggggctacctcaagg
aggacttcCTGgtcaagccagaggtgctgcagaaagccactgtggagctgctggaccaggcactgtgtgccagcttgt
acggccattcactcactgacaggatggtgtgcgctggctacctggacgggaaggtggactcctgccagggtgactcagg
aggacccctggtctgcgaggagccctctggccggttctttctggctggcatcgtgagctggggaatcgggtgtgcggaag
cccggcgtccaggggtctatgcccgagtcaccaggctacgtgactggatcctggaggccaccaccaaagcc

ctg22fin4 MOOSE14210 16664895..16664909, 16676917..16677063, 16677651..16677787, 16679898..16680169, 16681309..16681451
RIVGGAVSSEGEWPWQASLQVRGRHICGGALIADRWVITAAHCFQEDSMAS
TVLWTVFLGKVWQNSRWPGEVSFKVSRLLLHPYHEEDSHDYDVALLQLDH
PVVRSAAVRPVCLPARSHFFEPGLHCWITGWGALREGGPISNALQKVDVQLI
PQDLCSEVYRYQVTPRMLCAGYRKGKKDACQGDSGGPLVCKALSGRWFLA
GLVSWGLGCGRPNYFGVYTRITGVISWIQQTMAQS cgcattgttggtggagctgtgtcc
tccgagggtgagtggccatggcaggccagcctccaggttcggggtcgacacatctgtgggggggccctcatcgctgac
cgctgggtgataacagctgcccactgcttccaggaggacAGCatggcctccacggtgctgtggaccgtgttcctgggc
aaggtgtggcagaactcgcgctggcctggagaggtgtccttcaaggtgagccgcctgctcctgcacccgtaccacgaag
aggacagccatgactacgacgtggcgctgctgcagctcgaccacccggtggtgcgctcggccgccgtgcgccccgtct
gcctgcccgcgcgctcccacttcttcgagcccggcctgcactgctggattacgggctggggcgccttgcgcgagggcG
GCcccatcagcaacgctctgcagaaagtggatgtgcagttgatcccacaggacctgtgcagcgaggtctatcgctacca
ggtgacgccacgcatgctgtgtgccggctaccgcaagggcaagaaggatgcctgtcagggtgactcaggtggtccgct
ggtgtgcaaggcactcagtggccgctggttcctggcggggctggtcagctggggcctgggctgtggccggcctaactac
ttcggcgtctacacccgcatcacaggtgtgatcagctggatccagcaaaccatggcccagagc

ctg_2 MOOSE14212 2804144..2804283, 2805028..2805293, 2805383..2805519, 2805953..2806110, 2806154..2806157
RIVGGSAAGRGEWPWQVSLWLRRREHRCGAVLVAERWLLSAAHCFDVYG
DPKQWAAFLGTPFLSGAEGQLERVARIYKHPFYNLYTLDYDVALLELAGPV

RRSRLVRPICLPEPAPRPPDGTRCVITGWGSVREGGSMARQLQKAAVRLLSE
QTCRRFYPVQISSRMLCAGFPQGGVDSCSGDAGGPLACREPSGRWVLTGVT
SWGYGCGRPHFPGVYTRVAAVRGWIGQHIQDN aggattgtgggcggcagcgcag
cgggccgtggggagtggccgtggcaggtgagcctgtggctgcggcgccgggaacaccgttgcggggccgtgctggtg
gcagagaggtggctgctgtcggcggcgcactgcttcgacGTCtacggggaccccaagcagtgggcggccttcctag
gcacgccgttcctgagcggcgcggaggggcagctggagcgcgtggcgcgcatctacaagcacccgttctacaatctcta
cacgctcgactacgacgtggcgctgctggagctggcggggccggtgcgtcgcagccgcctggtgcgtcccatctgcct
gcccgagcccgcgccgcgacccccggacggcacgcgctgcgtcatcaccggctggggctcggtgcgcgaaggaGG
Ctccatggcgcggcagctgcagaaggcggccgtgcgcctcctcagcgagcagacctgccgccgcttctacccagtgca
gatcagcagccgcatgctgtgtgccggcttcccgcagggtggcgtggacagctgctcgggtgacgctgggggacccct
ggcctgcagggagccctctggacggtgggtgctaactggggtcactagctggggctatggctgtggccggccccacttc
ccaggtgtctatacccgggtggcagctgtgagaggctggataggacagcacatccagGACaac
ctgCHR3_11 MOOSE14225 15721072..15721214, 15722598..15722866,
15724464..15724600, 15726636..15726800
RIIGGTDTLEGGWPWQVSLHFVGSAYCGASVISREWLLSAAHCFHGNRLSDP
TPWTAHLGMYVQGNAKFVSPVRRIVVHEYYNSQTFDYDIALLQLSIAWPET
LKQLIQPICIPPTGQRVRSGEKCWVTGWGRRHEADNKGSLVLQQAEVELIDQ
TLCVSTYGIITSRMLCAGIMSGKRDACKGDSGGPLSCRRKSDGKWILTGIVS
WGHGSGRPNFPGVYTRVSNFVPWIHKYVPSL cgcatcatcggaggcacagacaccctgga
gggggttggccgtggcaggtcagcctccactttgttggatctgcctactgtggtgcctcagtcatctccagggagtggctt
ctttctgcagcccactgttttcatggaaacAGGctgtcagatcccacaccatggactgcacacctcgggatgtatgttcag
gggaatgccaagtttgtctccccggtgagaagaattgtggtccacgagtactataacagtcagacttttgattatgatattgct
ttgctacagctcagtattgcctggcctgagaccctgaaacagctcattcagccaatatgcattcctcccactggtcagagag
ttcgcagtggggagaagtgctgggtaactggctgggggcgaagacacgaagcaGATaataaaggctccctcgttctg
cagcaagcggaggtagagctcattgatcaaacgctctgtgtttccacctacgggatcatcacttctcggatgctctgtgcag
gcataatgtcaggcaagagagatgcctgcaaaggagattcgggtggacctttatcttgtcgaagaaaaagtgatggaaaat
ggattttgactggcattgttagctggggacatggaagtggacgaccaaactttcctggtgtttacacaagggtgtcaaacttt
gttccctggattcataaatatgtcccttctctt
ctg4012 MOOSE14231 9641523..9641665, 9645343..9645602,
9650813..9650955, 9656232..9656386, 9656688..9656691
RIVQGRETAMEGEWPWQASLQLIGSGHQCGASLISNTWLLTAAHCFWKNK
DPTQWIATFGATITPPAVKRNVRKIILHENYHRETNENDIALVQLSTGVEFSNI
VQRVCLPDSSIKLPPKTSVFVTGFGSIVDDGPIQNTLRQARVETISTDVCNRK
DVYDGLITPGMLCAGFMEGKIDACKGDSGGPLVYDNHDIWYIVGIVSWGQS
CALPKKPGVYTRVTKYRDWIASKTGMN agaattgtccaaggaagggaaacagctatg
gaaggggaatggccatggcaggccagcctccagctcatagggtcaggccatcagtgtggagccagcctcatcagtaac
acatggctgctcacagcagctcactgcttttggAAAaataaagacccaactcaatggattgctacttttggtgcaactata
acaccacccgcagtgaaacgaaatgtgaggaaaattattcttcatgagaattaccatagagaaacaaatgaaaatgacatt
gctttggttcagctctctactggagttgagtttcaaatatagtccagagagtttgcctcccagactcatctataaagttgccac
ctaaaacaagtgtgttcgtcacaggatttggatccattgtagatgatGGAcctatacaaaatacacttcggcaagccaga
gtggaaaccataagcactgatgtgtgtaacagaaaggatgtgtatgatggcctgataactccaggaatgttatgtgctggat
tcatggaaggaaaaatagatgcatgtaagggagattctggtggacctctggtttatgataatcatgacatctggtacattgta
ggtatagtaagttggggacaatcatgtgcacttcccaaaaaacctggagtctacaccagagtaactaagtatcgagattgg
attgcctcaaagactggtATGaac
ctg2019 MOOSE14237 67883182..67883313, 67883686..67883772,
67883860..67884019, 67884218..67884360, 67884574..67884741

RIVGGSAAPPGAWPWLVRLQLGGQPLCGGVLVAASWVLTAAHCFLLWTVT
LAEGSRGEQAEEVPVNRILPHPKFDPRTFHNDLALVQLWTPVSPGGSARPVC
LPQEPQEPPAGTACAIAGWGALFEDGPEAEAVREARVPLLSTDTCRRALGPG
LRPSTMLCAGYLAGGVDSCQGDSGGPLTCSEPGPRPREVLFGVTSWGDGCG
EPGKPGVYTRVAVFKDWLQEQMSGE cgcatcgtgggggcagcgcggcgccgcccggggc
ctggccctggctggtgaggctgcagctcggcgggcagcctctgtgcggcggcgtcctggtagcggcctcctgggtgctc
acggcagcgcactgctttcttctgtggactgtgacgctggcagaggggtcccgggggggagcaagcggaggaggtgcca
gtgaaccgcatcctgccccaccccaagtttgacccgcggaccttccacaacgacctggccctggtgcagctgtggacgc
cggtgagcccgggggggatcggcgcgccccgtgtgcctgccccaggagccccaggagccccctgccggaaccgcctg
cgccatcgcgggctggggcgccctcttcgaaGACgggcctgaggctgaagcagtgagagaggcccgtgttcccctg
ctcagcaccgacacctgccgaagagccctggggcccgggctgcgccccagcaccatgctctgcgccgggtacctggc
gggggggcgttgactcgtgccagggtgactcgggaggccccctgacctgttctgagcctggcccccgccctagagaggt
cctgttcggagtcacctcctgggggggacggctgcggggagccagggaagcccggggtctacacccgcgtggcagtgtt
caaggactggctccaggagcagatgagcggtgag
ctg4012 MOOSE14238 9538614..9538772, 9541857..9541999,
9546242..9546501, 9550024..9550160
RIASGVIAPKAAWPWQASLQYDNIHQCGATLISNTWLVTAAHCFQKYKNPH
QWTVSFGTKINPPLMKRNVRRFIIHEKYRSAAREYDIAVVQVSSRVTFSDDIR
QICLPEASASFQPNLTVHITGFGALYYGGESQNDLREARVKIISDDVCKQPQV
YGNDIKPGMFCAGYMEGIYDACRGDSGGPLVTRDLKDTWYLIGIVSWGDN
CGQKDKPGVYTQVTYYRNWIASKTGI agaatagcatctggagtcattgcacccaaggcggcct
ggccttggcaagcttcccttcagtatgataacatccatcagtgtggggccaccttgattagtaacacatggcttgtcactgca
gcacactgcttccagAAGtataaaaatccacatcaatggactgttagttttggaacaaaaatcaaccctcccttaatgaaa
agaaatgtcagaagatttattatccatgagaagtaccgctctgcagcaagagagtacgacattgctgttgtgcaggtctcttc
cagagtcacctttcggatgacatacgccagatttgtttgccagaagcctctgcatccttccaaccaaatttgactgtccacat
cacaggatttggagcactttactatggtGGGgaatcccaaaatgatctccgagaagccagagtgaaaatcataagtgat
gatgtctgcaagcaaccacaggtgtatggcaatgatataaaacctggaatgttctgtgccggatatatggaaggaatttatg
atgcctgcaggggtgattctgggggacctttagtcacaagggatctgaaagatacgtggtatctcattggaattgtaagctg
gggagataactgtggtcaaaaggacaagcctggagtctacacacaagtgacttattaccgaaactggattgcttcaaaaac
aggcatc
ctgCHR12_3 MOOSE14241 16346585..16346749, 16348623..16348765,
16374722..16374999, 16389747..16389901
RIIGGTEAQAGAWPWVVSLQIKYGRVLVHVCGGTLVRERWVLTAAHCTKD
ASDPLMWTAVIGTNNIHGRYPHTKKIKIKAIIIHPNFILESYVNDIALFHLKKA
VRYNDYIQPICLPFDVFQILDGNTKCFISGWGRTKEEGNYGNATNILQDAEV
HYISREMCNSERSYGGIIPNTSFCAGDEDGAFDTCRGDSGGPLMCYLPEYKR
FFVMGITSYGHGCGRRGFPGVYIGPSFYQKWLTEHFFHA cggattatgggggca
ccgaagcacaagctggcgcatggccgtgggtggtgagcctgcagattaaatatggccgtgttcttgttcatgtatgtgggg
gaaccctagtgagagagaggtgggtcctcacagctgcccactgcactaaagacgctAGCgatcctttaatgtggacag
ctgtgattggaactaataatatacatggacgctatcctcataccaagaagataaaaattaaagcaatcattattcatccaaact
tcattttggaatcttatgtaaatgatattgcacttttcacttaaaaaaagcagtgaggtataatgactatattcagcctatttgcct
acctttgatgttttccaaatcctggacggaaacacaaagtgttttataagtggctggggaagaacaaaagaagaaggtaatt
atGGTaacgctacaaatattttacaagatgcagaagtgcattatatttctcgagagatgtgtaattctgagaggagttatgg
gggaataattcctaacacttcattttgtgcaggtgatgaagatggagcttttgatacttgcaggggtgacagtgggggaccat
taatgtgctacttaccagaatataaaagatttttttgtaatgggaattaccagttacggacatggctgtggtcgaagaggttttcc
tggtgtctatattgggccatccttctaccaaaagtggctgacagagcatttcttccatgca

ctgCHR12_2 MOOSE14247 26347557..26347748, 26349485..26349636, 26350962..26351057, 26355150..26355322, 26356046..26356143, 26356421..26356468 RISSWRNSTVTGHPWQVSLKSDEHHFCGGSLIQEDRVV TAAHCLDSLSEKQLKNITVTSGEYSLFQKDKQEQNIPVSKIITHPEYNSREYM SPDIALLYLKHKVKFGNAVQPICLPDSDDKVEPGILCLSSGWGKISKTSEYSN VLQEMELPIMDDRACNTVLKSMNLPPLGRTMLCAGFPDWGMDACQGDSG GPLVCRRGGGIWILAGITSWVAGCAGGSVPVRNNHVKASLGIFSKVSELMDF ITQNLFTG (SEQ ID NO: 79)

agaattagtagttggagaaattcaacagtgactggacatccatggcaggtctccctaaaatcagatgagcaccacttctgtg gaggaagcttgattcaagaagatcgggttgttacagcagcacactgcctggacagcctcagtGAGaagcagctgaag aatataactgtgacttctggggagtacagcctctttcagaaggataagcaagaacagaatattcctgtctcaaaaattattac ccatcctgaatacaacagccgtgaatatatgagtcctgatattgcactgctgtatctaaaacacaaagtcaagtttGGAaat gctgttcagccaatctgtcttcctgacagcgatgataaagttgaaccaggaattctttgcttatccagtggatggggcaagat ttccaaaACAtcagaatattcaaatgtcctacaagaaatggaacttcccatcatggatgacagagcgtgtaatactgtgct caagagcatgaacctccctccccctgggaaggaccatgctgtgtgctggcttccctgattggggaatggacgcctgccagg gggactctggaggaccactggtttgtagaagaggtggtggaatctggattcttgctgggataacttcctgggtagctggttg tgctggaggttcagttcccgtaagaaacaaccatgtgaaggcatcacttggcattttctccaaagtgtctgagttgatggattt tatcactcaaaacctgttcacaggt (SEQ ID NO: 32)

ctg4012 MOOSE14251 10140958..10140985, 10163010..10163036, 10189245..10189369, 10190482..10190601, 10191060..10191348, 10192103..10192190, 10195165..10195210 EIWSGEQGQNDMVWLSSLKMSGQHYCGASLISERHLVTAAHCFKVTKNPK NYTVSFGTKVTLPYMQHDVQQIIIHEDYIQDEHHDDIALILLTKKVLFKNDVH RVCLPEATQIFPPGEGVVVTGWGRLSFNGKISENLTYHKASVKITDTNTCNA KEAYRSMVQDRVLCAGYMEGNIDACQGDSGGPLVHPNSLNIWYIWYLVGV VSWGRNECGAINSPGVYTQTDVFFFLKWIKSTIALK (SEQ ID NO: 80)

gagatttggtcaggggaacaggggcagaatgatatggtttggctcTCAagccttaaaatgagtgggcaacactactgtg gggcatcattgatcagtgaaagacacttggtgactgcagctcactgttttaaaGTGacaaaaaatccaaaaaactatact gtcagctttggcacgaaagtaactcttccctatatgcaacatgatgttcaacaaattattattcatgaagactacatccaggat gaacatcatgatgatatcgcacttatactgctcactaaaaaagttttatttaagaatgatgtacatcgagtttgtcttcctgaagc cacacagatttttccacctggtgaaggagttgttgttacaggatggggaagactttcatttaatggtaagatcagtgaaaactt aacataccataaagcatctgtgaagattactgatacaaacacttgtaatgctaaagaagcctatcgtagtatggtacaggata gagtgctatgtgctgggtacatggaaggaaatatagacgcctgccagggagactctggaggaccactagttcatcctaatt ctctaaatatttggtatatttggtaccttgttggagtagtgagctggggaaggaatgaatgtggtgcaatcaatagtccaggg gtctacACAcagacagatgtctttttttttttaAAGtggatcaaaagcacaattgctctcaaa (SEQ ID NO: 33)

ctg4015 MOOSE14254 14157594..14157630, 14167829..14167892, 14178504..14178769, 14179473..14179642, 14187478..14187633 RIVSMESKKGKVQWLVVLFGSSSIQGSRKDKAIKTWTTFSYTVWLGSITVGD SRKRVKYYVSKIVIHPKYQDTTADVALLKLSSQVTFTSAILPICLPSVTKQLAI PPFCWVTGWGKVKESSDRDYHSALQEAEVPIIDRQACEQLYNPIGIFLPALEP VIKEDKICAGDTQNMKDSCKGDSGGPLSCHIDGVWIQTGVVSWGLECGKSL PGVYTNVIYYQKWINATISRA (SEQ ID NO: 81)

aggatagtcagcatggaatctaagaagggaaaagtcCAAtggctagtggtcctgtttggcagctcttccattcagggaa gcaggaaagataaggccataAAGacctggactacttttttcatatactgtgtggctaggatcgattacagtaggtgactca aggaaacgtgtgaagtactacgtgtccaaaatcgtcatccatcccaagtaccaagatacaacggcagacgtcgccttgttg aaactgtcctctcaagtcaccttcacttctgccatcctgcctatttgcttgcccagtgtcacaaagcagttggcaattccaccct

tttgttgggtgaccggatggggaaaagttaaggaaagttcaGATagagattaccattctgcccttcaggaagcagaagt
acccattattgaccgccaggcttgtgaacagctctacaatcccatcggtatcttcttgccagcactggagccagtcatcaag
gaagacaagatttgtgctggtgatactcaaaacatgaaggatagttgcaagggtgattctggagggcctctgtcgtgtcaca
ttgatggtgtatggatccagacaggagtagtaagctggggattagaatgtggtaaatctcttcctggagtctacaccaatgt
aatctactaccaaaaatggattaatgccactatttcaagagcc (SEQ ID NO: 34)
ctgchr7_ctg18037 MOOSE14260 3060956..3061102, 3061204..3061343,
3063795..3064042, 3064266..3064399
LAFNPDYTVSSTPPYLVYLKSDYLPCAGVLIHPLWVITAAHCNLPKLRVILGV
TIPADSNEKHLQVIGYEKMIHHPHFSVTSIDHDIMLIKLKTEAELNDYVKLAN
LPYQTISENTMCSVSTWSYNVYKEPDSLQTVNISVISKPQCRDAYKTYNITEN
MLCVGIVPGRRQPCKEVSAAPAICNGMLQGILSFADGCVLRADVGIYAKIFY
YIPWIENVIQNN (SEQ ID NO: 82)
ttggcctttaatccagattacacagtcagctccactccccttacttggtctatttgaaatctgactacttgccctgcgctggag
tcctgatccacccgctttgggtgatcacagctgcacactgcaatttaCCAaagcttcgggtgatattgggggttacaatcc
cagcagactctaatgaaaagcatctgcaagtgattggctatgagaagatgattcatcatccacacttctcagtcacttctattg
atcatgacatcatgctaatcaagctgaaaacagaggctgaactcaatgactatgtgaaattagccaacctgccctaccaaa
ctatctctgaaaataccatgtgctctgtctctacctggagctacaatgtgTACaaagagcccgattcactgcaaactgtga
acatctctgtaatctccaagcctcagtgtcgcgatgcctataaaacctacaacatcacggaaaatatgctgtgtgtgggcatt
gtgccaggaaggaggcagccctgcaaggaagtttctgctgccccggcaatctgcaatgggatgcttcaaggaatcctgtc
ttttgcggatggatgtgttttgagagccgatgttggcatctatgccaaaatttttttactatataccctggattgaaaatgtaatcc
aaaataac (SEQ ID NO: 35)
ctg2014 MOOSE14265 21780003..21780336, 21782470..21782645,
21786391..21786549 RWAAGVRVPAQHSEEPPHNRSTNPSDYRILLGYDQQS
HPTEHSKQMTVNKIMVHADYNELHRMGSDITLLQLHHHVEFSSHILPACLPE
PTTWLAPDSSCWISGWGMVTEDVFLPEPFQLQEAEVGVMDNTVCGSFFQPQ
YPGQPSSSDYTIHEDMLCAGDLITGKAICRRDSRGPLVCPLNGTWFLMGLSS
WSLDCCSPVGPRVFTRLPYFTNWISQKKRES (SEQ ID NO: 83)
aggtgggcagccggggtgagggtgccggcccagcattcagaggagcctccccacaacaggtccactaacccatctgat
taccggatcctgcttgggtatgaccagcaaagccatcccacagagcacagcaagcagatgacagtgaataagatcatggt
gcacgctgactataacgagttgcaccgcatggggagtgacatcaccctgctgcagctgcaccatcatgtggaattcagctc
ccacatcctccccgcctgccttccggaaccaaccacgtggctggcccctgacagctcctgctggatatctggttggggaat
ggtcaccgaggatGTCttcctgcctgagcccttccaacttcaggaggcagaggtcggtgtcatggacaacactgtctgc
ggatccttttccagccccagtaccccggccagccaagcagcagtgactacaccatccacgaggacatgctgtgcgctgg
ggacctcataacaggaaaggccatttgccgacgagactccaggggtcccctcgtctgcccattaaatggcacctggttcct
gatggggctgtctagttggagcctcgactgctgctcacccgtcggtcccagggtcttcaccaggctcccctacttcaccaa
ctggatcagccagaagaagagggagagc (SEQ ID NO: 36)
ctgC20 MOOSE14272 1992244..1992402, 1993117..1993202,
1996233..1996286, 2017918..2017988, 2050580..2050652, 2054829..2054897,
2058461..2058508, 2059496..2059523, 2064130..2064150
RVVSGYFSANMVSTPWRTGILHFNHCIHDLSQTVLGDHLVKFHHTIKIICHIL
DHAVALLFLQISSIWNGNIYPIPLPAFVSYKNASICRIMLWGHAGDMLFPMNF
PLCARVDRQQGEQCEHTEFGYQPETIKNDMLCAGFEEGKKDACKGDSGGPL
VCLVGQSWLQAGVISWGEGCARQNRPGVYIRVTAHHNWIHRIIPKL (SEQ
ID NO: 84)
agagttgtttctggatacttttcagcaaacatggtttctactccctggAGAacaggcattttacattttaaccactgcattcat
gatctgagccaaACAgtcctgggggatcatttagttaaattccatcatactataaagattatttgccatatattagatcatG

CTgtggccctttgtttttgcaaatttcttccatttggaatgggaacatttacccaatacctctacctgcaTTTgtttcctaca
agaatgctagtatttgtaggatcatgttgtggggacatgctggggacatgcttttccccATGaactttcccttgtgtgcccg
cgtggacagacaacagggggagcagtgcgagCACaccgagtttggctaccaacccgaaaccatcaagaatgacatg
ctgtgcgccggcttcgaggagggcaagaaggatgcctgcaagggcgactcgggcggccccctggtgtgcctcgtgggt
cagtcgtggctgcaggcgggggtgatcagctggggtgagggctgtgcccgccagaaccgcccaggtgtctacatccgt
gtcaccgcccaccacaactggatccatcggatcatccccaaactg (SEQ ID NO: 37)
ctgChr_1ctg109 MOOSE14278 4832..4921, 5738..5839, 6532..6615,
6661..6758, 10595..10623, 15756..15898, 16240..16398
HIINGKRQIAFPRRPGTREGCPLLLFLSNAHCTPPWATEQDSNSKKKKKKETE
KTIPKATVIKTDGHYKENKNRKHQVLAKMWRNWNLYALLVFCKIKHRITEP
GRVAHACNPSTLGGRGGWITRWGSHYVAQAGETSDELQEMQLPLILEPWC
HLLYGHMSYIMPDMLCAGDILNAKTVCEGDSGGPLVCEFNRSWLQIGIVSW
GRGCSNPLYPGVYASVSYFSKWICDNIEIT (SEQ ID NO: 85)
catataatcaatggtaaaagacagatagctttcccccgaagaccaggaacaagagaaggatgtccacttttgctatttctatc
caatgcacactgcactccgccatgggcaacagagcaagactccaactcaaaaaaaaaaaaaaaaaaagagacagaga
aaacaattccaaaagctacagttatcaaaacagatggccactataaagaaaacaaaaacagaaaacatcaagtgttggca
aagatgtggagaaattggaacctttatgcactgttggttttctgcaagattaaacatagaattactgagccaggcagggtggc
tcacgcctgtaatcccagcactttgggaggccgaggcgggtggatcacgAGAtgggggtctcactatgttgcccaggc
tGGTgagacctcagacgagctgcaggagatgcagctcccgctgatcctggagccctggtgccacctgctctacggac
acatgtcctacatcatgcccgacatgctgtgtgctggggacatcctgaatgctaagaccgtgtgtgagggcgactccggg
ggcccacttgtctgtgaattcaaccgcagctggttgcagattggaattgtgagctggggccgaggctgctccaaccctctgt
accctggagtgtatgccagtgtttcctatttctcaaaatggatatgtgataacatagaaatcacg (SEQ ID NO: 38)
ctgChr_6ctg20 MOOSE14279 3772814..3772839, 3773615..3773663,
3774882..3774902, 3776288..3776466, 3786503..3786609, 3803444..3803462,
3811761..3811774, 3825365..3825494, 3826636..3826668, 3842185..3842229,
3846239..3846299 RVSGGRDSVPSLVPSTNAYNRKRPENPHMCGGFLASNI
EHLLCARHRIQKSMTSAHRSKVRRLESHWYKGKRKTRSKEKRKIFGKYTSNI
NYDISLLGLASPAVITDKVIPACLPSPNYVVADQTECYITDWGETQGTFGAGF
LKEAQLPVIENEVCNRYEFLNGRVKSTELCAGHLAGGIDSCKVRKDQETKVS
LFGIGCGDWVRSPHFYTYIHTYTPSIQENIKEN (SEQ ID NO: 86)
agggtctctggaggtagggacagtgtcccatctttggtaccatccaccaatgcctacaacAGGaagaggcctgagaac
cctcacatgtgtggaggtttcctggccTCAaacattgagcacctgctgtgtgctaggcacAGGattcaaaaatccatg
acgtctgctcataggtcaaaggttaggagacttgaatctcattggtacaaagggaaaagaaagacaaggagtaaagagaa
aaggaaaatatttggaaaatacaccagcaacATAaattacgacataAGTctgctgggtttggccAGTcctgccgtc
atcactgacaaagtaatcccagcttgtctgccatccccaaattatgtggtcgccgaccagactgaatgttacatcactgactg
gggagaaacccaaGGTacctttggggctggctttctcaaggaagcccagctccctgtgattgagaatgaagtgtgcaat
cgctatgagtttctgaatggaagagtcaaatccactgagctctgtgctgggcatttggctggaggcattgacagttgcaagg
taagaaaagatcaagagaccaaagttagtctttttggtataggatgtggagattgggttaggtccccacatttttatacataca
tacacacatacacaccgTCCattcaagaaaatatcaaagaaaat (SEQ ID NO: 39)
ctg11ctg15 MOOSE14295 6743906..6743990, 6796305..6796757,
6824377..6824594 RKLGILNHQVLFWYNLSLLLHFIGYKSYSEPLALFGEDD
DMDPRPSRSYQVANGIAVLPVSGTLVSKTRALQPYSGMTGYNGIIARLQQAI
SDPGVDGILLDMDTPGGMVSGAFDCADIIARMRDIKPIWALANDMNCSAGQ
LIASSASRRLVTQTARTGSIGVMMAHSNYGAALKTNGGHMHTYVYCSTIHN
SKDLKPTQMPINNRLDKENVAHIHHGILCSHKKDEFMSFAGTWMKLETIILS
KLTQEQK (SEQ ID NO: 87)

agaaagctaggcatcttaaaccaccaggtactattttggtataatctatcacttctgttacattttattggatataaatcatatTC
Cgaaccgctggcgctgtttggtgaggatgatgacatggatccccgtccatcacgcagctatcaggtggcaaatggtatcg
cggtcttgccggtttccggcacgctggtcagtaaaacccgtgcgcttcagccttattccgggatgacgggttacaacggga
tcattgctcgcctgcagcaggcaatcagtgaccccggcgttgacggcattcttctggatatggatacgccgggtggaatgg
tgtccggggcgtttgactgcgccgacattattgcccgtatgcgcgatatcaaacccatctgggcgctggccaatgacatga
actgcagtgcaggtcagcttattgccagttcggcatcgcgacggctggtcacacaaacggccagaaccggctccattggg
gtcatgatggcgcacagtaactatggcgctgcgctcaaaactaacggcGGGcacatgcacacatatgtttattgcagca
ctattcacaatagcaaagacttaaaaccaacccaaatgcccatcaataatagactggataaagaaaatgtggcacatatac
accatggaatactgtgcagtcataaaaaggatgagttcatgtcctttgcaggaacgtggatgaagctagaaaccatcattct
cagcaaactaacacaggaacagaaa (SEQ ID NO: 40)

ctgC31 MOOSE14296 605248..605324, 613146..613186, 625648..625692, 631870..631946, 638586..639242

MLGVLLQIWRGSWKKQTQAQGRRERSRQAAGAVSAGGRRALLLYLRAELE
DKLACVDSRLRLVMRGLVLGRASGSSVRPKLPKDVRADFQTRIDATRQMFA
EKVSAYTGMSVQDVLDTEAAVFSGQESLDNGLADELVNNTDALGVMREAL
DRRKKTTLGGTMPSPSASAVTTKPVDQAATQTTASAEQATTVDTTIASVAA
PVDVSAQVTAAVAAENSRIMGILNCDEAKGRESQARALAETPGMTVESAQR
ILAAAPQSAQMRTDTALDRLMETAPGALQAGSASSDAADDLLNTPV (SEQ ID NO: 88)

atgctgggggtgcttcttcaaatctggaggggagctggaagaagcagacacaggcccagggccggagggagagaA
GCcgccaggcagccggggctgtgagtgccggaggccgacgcGCActgttattgtatcttagagctgaactggaaga
caaactggccTGTgtggacagcaggcttagactggtgatgaggggcttgtcctggggagggcctcaggcagctctg
ttagacccaaacttccgaaagatgtgcgtgctgatttccagacgcgtatcgatgccactcgtcagatgtttgccgaaaaggtt
tccgcttataccggcatgtctgttcaggacgtgctggacaccgaagcggcagtattctccggccaggaatctttggataac
gggctggcggatgaacttgttaacaataccgatgcgctcggcgtgatgcgcgaagcactcgacagacgcaaaaaaacaa
cccttggaggaactatgccatcaccttctgcatcagctgtgaccactaagccagttgaccaggcagcaactcagacaactg
catcagctgaacaggccactaccgttgacacgacaattgcttccgtagcagcccctgtagatgtcagtgcgcaggttactg
cagcagtagctgcagagaatagtcgcatcatgggcatcctgaactgcgacgaggctaaagggcgtgagtcacaggcgc
gagcactggccgaaacgccgggtatgacggtagagagcgcacagcgcattctggctgctgcaccgcaaagtgcccag
atgcgtaccgatacggcgctggatcgtttgatggaaacagcacccggtgcactccaagcaggtagcgcatcttctgatgc
cgctgacgatttgttaaacacccccgtt (SEQ ID NO: 41)

ctg_2 MOOSE14301 1416405..1416576, 1416840..1417019, 1417206..1417323, 1417394..1417497, 1417583..1417683, 1420409..1420621, 1420697..1420869, 1421159..1421247, 1421341..1421417, 1421520..1421648, 1421744..1421776

TDPWFSKQWYMNSEAQPDLSILQAWSQGLSGQGIVVSVLDDGIEKDHPDLW
ANYDPLASYDFNDYDPDPQPRYTPSKENRHGTRCAGEVAAMANNGFCGVG
VAFNARIGGVRMLDGTITDVIEAQSLSLQPQHIHIYSASWGPEDDGRTVDGP
GILTREAFRRGVTKGRGGLGTLFIWASGNGGLHYDNCNCDGYTNSIHTLSV
GSTTQQGRVPWYSEACASTLTTTYSSGVATDPQIVTTDLHHGCTDQHTGTS
ASAPLAAGMIALALEANPFLTWRDMQHLVVRASKPAHLQAEDWRTNGVGR
QVSHHYGYGLLDAGLLVDTARTWLPTQPQRKCAVRVQSRPTPILPLIYIREN
VSACAGLHNSIRSLEHVQAQLTLSYSRRGDLEISLTSPMGTRSTLVAIRPLDV
STEGYNNWVFMSTHFWDENPQGVWTLGLENKGYYFNTGEGGAGLWWAG
LGSPT (SEQ ID NO: 89)

acggacccctggttctccaagcagtggtacatgaacagcgaggcccaaccagacctgagcatcctgcaggcctggagtc
aggggctgtcaggccagggcatcgtggtctctgtgctggacgatggcatcgagaaggaccacccggacctctgggcca

actacgacccccctggccagctatgacttcaatgactacgacccggaccccccagccccgctacacccccagcaaagaga
acCGGcacgggacccgctgtgctggggaggtggccgcgatggccaacaatggcttctgtggtgtgggggtcgctttc
aacgcccgaatcggaGGCgtacggatgctggacggtaccatcaccgatgtcatcgaggcccagtcgctgagcctgca
gccgcagcacatccacatttacagcgccagctggggtcccgaggacgacggccgcacggtggacggccccggcatcc
tcacccgcgaggccttccggcgtggtgtgaccaagggccgcggcgggctgggcacgctcttcatctgggcctcgggca
acggcggcctgcactacgacaactgcaactgcgacggctacaccaacagcatccacacgctttccgtgggcagcacca
cccagcagggccgcgtgccctggtacagcgaagcctgcgcctccaccctcaccaccacctacagcagcggcgtggcc
accgacccccagatcgtcaccacggacctgcatcacgggtgcacagaccagcacacgggcacctcggcctcagcccc
actggcggccggcatgatcgccctagcgctggaggccAACccgttcctgacgtggagagacatgcagcacctggtg
gtccgcgcgtccaagccggcgcacctgcaggccgaggactggaggaccaacggcgtggggcgccaaGTGagcca
tcactacggatacgggctgctggacgccgggctgctggtggacaccgcccgcacctggctgcccacccagccgcaga
ggaagtgcgccgtccgggtccagagccgccccACCcccatcctgccgctgatctacatcagggaaaacgtatcggcc
tgcgccggcctccacaactccatccgctcgctggagcacgtgcaggcgcagctgacgctgtcctacagccggcgcgga
gacctggagatctcgctcaccagccccatgggcacgcgctccacactcgtggccataCGAcccttggacgtcagcact
gaaggctacaacaactgggtcttcatgtccacccacttctgggatgagaacccacagggcgtgtggaccctgggcctaga
gaacaagggctactatttcaacacgggtgagggcggggcggggctgtggtgggcggggcttggctctccaacc
(SEQ ID NO: 42)
ctg18ctg2 MOOSE14331 595090..595301, 599634..599677, 601386..601450,
601725..601849, 609022..609148, 619640..619741
MASRYDRAITVFSPDGHLFQVEYAQEAVKKGSTAVGIRGTNIVVLGVEKKS
VAKLQDERTVRKICALDDHVCMAFAGLTADARVVINRARVECQSHKLTVE
DPVTVEYITRFIATLKQINTKSYLKFSREVPFLFCFLFFSWDYRHMPPHLANFF
AGYKINKQKFAAFLYANNEQSEKEIKKVIPFMIATNKIKCIEINLTKEVKDFH
NENYKTLMQETEADTKK (SEQ ID NO: 90)
atggcgtctcgatatgacagggcgatcactgtcttctccccagacggacacctttttcaagttgaatatgcccaggaagcgg
tgaagaaaggatccaccgcggtcggaattcgaggtaccaatatagttgttcttggggtagaaaaaaaatctgttgccaagct
tcaagatgaaagaactgtgaggaaaatttgtgcccttgatgaccatgtctgcatggcttttgcaGGActtactgctgatgct
agagtagtaataaacagagcccgtgtggagtgccagagccataagcttacggttgaggacccagtcactgtagaatacat
aactcgcttcatagcaactttaaagcagattaatacaaagagttatttgaagttttccagagaagtacctttttgttttgttttttgt
ttTTTagctgggattaccggcacatgccaccacacctggctaactttTTTgcaggatacaaaatcaacaaacaaaaatt
tgcagcatttctatatgccaacaatgaacaatctgaaaaagaaatcaagaaagtaatcccatttatgatagctacaaataaaa
ttaaatgcatagaaataaacttaaccaaagaagtgaaagatttccacaatgaaaactataaaacactgatgcaagaaactga
agcagacaccaaaaaa (SEQ ID NO: 43)
ctg18ctg2 MOOSE14335 803151..803182, 814164..814218, 819027..819098,
831305..831344, 838218..838243, 839545..839577, 840006..840128,
860714..860833, 860952..861014, 887374..887493
SKGGISVGLCVRDGVVVVSRDTNSPHRVTPLLNELMCLRCSGLAAAAAKMV
AAFISLRRSAEINKYVIYPRDVCTPYIVNRMSLIKIKYTQSNGRRPFGISALIVG
FDDDGISRLYQTDPSGTYHAWKANAIGRSAKTVREFLEKNYTEDAIASDSEA
IKLAIKALLEVVQSGGKNIELAIIRRNQPLKKKEEEEERRKKKEEEGGEEEEEE
EEEDEEEEEEVEEEEEE (SEQ ID NO: 91)
tcaaaaggaggaatttcagtgggtctctgtGTTcgggatggggtggtggtggtaagtagagatactaacagccctcaca
gagttactcctctgctaaatgaactaatgtgtctcaggtgttctgggctggcagcagctgcgaagatggttgcagcattcatc
tctctgaggagatcagcagagataaataagtatgttATAtatccaagagatgtatgcacccctatatagtgaacagaatg
tccttgataaaaataaaatatacccaaagcaatggacgaagacctttggtatttctgccttaattgtaggttttgatgatgatg
gtatctcaagattgtatcagacagatccttctggtacttatcatgcttggaaggcaaatgcaataggccgaagtgctaaaact

gttcgagaatttctagaaaagaattacacagaagatgccatagcaagtgacagtgaagctatcaagttagcaataaaagctt
tgctagaagttgtccagtctggtggaaaaaacattgaacttgctataataagaagaaatcaacctttgaagaagaaagagga
ggaggaggaaagaaggaagaagaaagaagaagaaggaggagaagaagaagaggaggaggaggaagaggatgag
gaggaagaggaggaggtggaagaggaagaggaagaa (SEQ ID NO: 44)

ctgChr_9ctg1926 MOOSE14337 2164211..2164258, 2164700..2167526, 2169222..2169361 ENGSLTWQELLRQTGKCSIPCLIDTGAQANIITEETVRA HKLPTRPWSKSVIYGGVYPNKINRKTIKLNISLNGISIKTEFLVVKKFSHPAAIS FTTLYDNNIEISSSKHTLSQMNKVSNIVKEPELPDIYKEFKDITAETNTEKLPK PIKGLEFEVELTQENYRLPIRNYPLPPGKMQAMNDEINQGLKSGIIRESKAINA CPVMFVPKKEGTLRMVVDYKPLNKYVKPNIYPLPLIEQLLAKIQGSTIFTKLD LKSAYHLIRVRKGDEHKLAFRCPRGVFEYLVMPYGISIAPAHFQYFINTILGE VKESHVVCYMDNILIHSKSESEHVKHVKDVLQKLKNANLIINQAKCEFHQSQ VKFIGYHISEKGFTPCQENIDKVLQWKQPKNRKELRQFLGSVNYLRKFIPKTS QLTHPLNNLLKKDVRWKWTPTQTFQAIENIKQCLVSPPVLRHFDFSKKILLET DASDVAVGAVLSQKHDDDKYYPVGYYSAKMSKAQLNYSVSDKEMLAIIKS LKHWRHYLESTIEPFKILTDHRNLIGRITNESEPENKRLARWQLFLQDFNFEIN YRPGSANHIADALSRIVDETEPIPKDSEDNSINFVNQISITDDFKNQVVTEYTN DTKLLNLLNNEDKRVEENIQLKDGLLINSKDQILLPNDTQLTRTIIKKYHEEG KLIHPGIELLTNIILRRFTWKGIRKQIQEYVQNCHTCQINKSRNHKPYGPLQPIP PSERPWESLSMDFITALPESSGYNALFVVVDRFSKMAILVPCTKSITAEQTAR MFDQRVIAYFGNPKEIIADNDHIFTSQTWKDFAHKYNFVMKFSLPYRPQTDG QTERTNQTVEKLLRCVCSTHPNTWVDHISLVQQSYNNAIHSATQMTPFEIVH RYSPALSPLELPSFSDKTDENSQETIQVFQTVKEHLNTNNIKMKKYFDMKIQE IEEFQPGDLVMVKRTKTAFLYTNNRQTESQIMSELPFTIASKRIKYLGIQLTRE VKDLFKENYK (SEQ ID NO: 92)

gagaatggaagtcttacatggcaagaattactcagacaaacagggaaatgctcaataccctgtctaatcgatacgggcgct
caagcaaatattataacagaagaaactgttcgagcacataaactgcctaccagaccctggtcaaaaagtgtgatatatggtg
gagtttatccaaataagattaatcgcaaaacaataaaacttaacataagtctaaatggaatatcaatcaaaacagaattcttgg
ttgtaaagaaatttcgcatccagctgctatctccttcacaacattatatgacaataacattgaaatatctagcagtaaacacac
gctctctcaaatgaacaaagtttcaaatattgtcaaggaacctgagttaccagatatctataaagaattcaaagacattactgc
agaaaccaatacggaaaagctaccaaagccaataaaagggttagaatttgaagttgaactaactcaagaaaactacagatt
acctatcagaaattacccgctaccaccgggaaaaatgcaagctatgaatgatgaaattaatcaaggattaaaaagtggaatt
atacgagaatctaaagccattaacgcctgtccagtaatgttcgttccgaaaaaggaaggcaccttgagaatggtggttgact
acaaaccttaaataagtatgtcaaacccaatatatatccgttaccacttattgaacaattacttgctaaaatacaaggttctac
aatttttactaaacttgacctcaaaagtgcctatcacttgatacgagtaagaaaaggagatgaacataaacttgcttttcgctgt
cctcgtggagttttgaatatctagtaatgccttatggcatatctatagctccagcacattttcaatactttatcaatacaatactt
ggtgaagtcaaagaatcgcatgtagtatgttatatggataatattttaattcattcaaaatcggaatctgaacatgtaaaacatg
ttaaagacgttctacagaaattaaaaaatgcgaacttaattatcaatcaagcaaaatgtgaatttcatcaatcacaagtaaaat
ttatagggtatcacatttcggaaaaaggatttacgccttgtcaagaaaatatagacaaagtcttacaatggaagcaacctaag
aatcgtaaagaattacgacaatttctaggttctgtcaattatcttaggaaattcattccaaagacatcacaattaacacatccac
tcaataatcttttgaaaaaggatgtacgctggaaatggacaccaacacaaacccaagcgatagaaaacattaaacaatgttt
agtttcgcctccggtgctacgacactttgatttcagtaaaaagattctactagaaactgatgcttcagatgtcgctgtaggagc
cgtattgtctcaaaaacatgatgatgataaatactatcctgttggatactattcagcaaagatgtctaaagcacaattaaattat
agcgtatcggacaaagaaatgcttgcaatcattaagtctctcaaacattggagacactatttagaatccactatcgaacctttc
aaaattttaacagaccatcgaaacttaattggtcgcattactaacgaatccgagcctgaaaacaaacgtttagctcgttggca
attatttttacaagacttcaactttgaaattaactacagacctggatcagcaaatcacatagctgatgccttatccagaattgttg

acgaaacagaaccaattccaaaagattcagaagacaatagtatcaactttgttaatcaaatctcgataaccgatgattttaaa
aaccaagtggttacagaatatacgaatgatacaaaattgttgaatttactaaacaatgaagacaaacgagtggaagagaata
tccaactcaaagatggcttactaattaacagtaaagaccaaatcttattacctaatgatactcagctgactaggacaattatta
aaaagtatcatgaagaaggtaaattgattcatccaggcattgaacttcttacaaacattatattacgtagatttacgtggaaag
gaataagaaaacaaatacaagaatatgtacagaactgccatacatgtcaaataaacaaatctaggaatcataaaccttatgg
acctttacaaccaattcccccatcagaaagaccttgggaatctttatcaatggattttattacagctttaccagaatcatctggtt
ataatgcacttttcgtggtagttgaccgattttcaaaaatggcaatcttagtaccttgtacgaaatccattacagcagagcaaa
cagctcgaatgtttgatcaacgagttattgcttatttcggcaatccaaaagaaatcattgcagataatgatcatatttttacttcc
caaacgtggaaagatttcgcacataaatataatttcgttatgaaattttcgttaccatacagaccacaaactgatggacaaact
gagcgtacaaaccaaactgtggagaaattactaagatgtgtatgtagcacacatccaaatacatgggtagatcatatatccc
tagtgcaacaatcttacaacaatgcgatacattcagcaactcaaatgacacctttgagatagtacatcgctattcaccagctt
tatcacctttagagttacctagctttagtgacaaaactgacgaaaactctcaggaaacgatccaagtatttcaaacagttaaa
gaacacttgaatacaaacaacataaagatgaaaaagtatttcgatatgaaaatacaagaaattgaagaatttcaacctggag
acctagttatggtcaaaagaacgaaaacaGCAttcttatacaccaataacagacaaacagagagccaaatcatgagtga
actcccattcacaattgcttcaaagagaataaaatacctaggaattcaactgacaagggaagtgaaggacctcttcaagga
gaactacaaa

ctgchr14_1 MOOSE14342 3984761..3985345

GPRLAHGTTTLAFRFRHGVIAAADTRSSCGSYVACPASCKVIPVHQHLLGTT
SGTSADCATWYRVLQRELRLRELREGQLPSVASAAKLLSAMMSQYRGLDL
CVATALCGWDRSGPELFYVYSDGTRLQGDIFSVGSGSPYAYGVLDRGYRYD
MSTQEAYALARCAVAHATHRDAYSGGSVDLFHVRESGWEHVS ggccccag
actggcccacggcaccaccactctggccttccgcttccgtcatggagtcattgctgcagctgacacgcgttcctcctgtgg
cagctatgtggcgtgtccagcctcatgcaaggtcatccctgtgcaccagcacctcctgggtaccacctctggcacctctgc
cgactgtgctacctggtatcgggtattacagcgggagctgcggcttcgggaactgagggagggtcagctgcccagtgtg
gccagtgctgccaagctcttgtcagccatgatgtctcaataccggggactggatctctgtgtggccactgccctctgcggct
gggaccgctctggccctgagctcttctacgtctatagcgacggcacccgcctgcaggggacatcttctctgtgggctctg
gatctccctatgcctacggcgtgctagaccgtggctatcgctacgacatgagcacccaggaagcctacgccctggctcgc
tgcgccgtggcccacgccacccaccgtgatgcctattcaggggctctgtagacctttccacgtgcgggagagtggatg
ggagcatgtgtca

ctg2019 MOOSE14348 50831031..50831624

SIMSYNGGAIMAMKGKNRVAIAADRHFGIQAQMVTTDFQEIFPMGGWLYIG
LAGLATDVQRVAQCLKFQLNLYELKEGQQIKPYTFTSMVANFLYEKHFGPY
YTDPVIAGLDLKTFKPFSCSLDLIGFPMVTDDFVVNGSYAEQMYGMCESLW
EPNMDPEHPFETISPAMLNAVDWGAGSGMGVIIHITKKDKITTRT tctattatgtc
ctataacggaggagccatcatggccatgaaggggaagaaccgtgtggccatcgctgcagacaggcacttcgggatcca
ggcccagatggtgaccacggacttccaggagatctttcccatgggtggttggttgtacatcggtctggccgggcttgccac
tgacgtccagagagttgcccagtgcctcaagttccagctgaacctatatgagttgaaggaaggtcagcagatcaaacctta
taccttcacgagcatggtggccaacttcttgtatgagaaacattttggcccctactacactgatccagtcattgctggtttgga
cctgaagacctttaagcccttcagttgctctctagacctcatcggcttccccatggtgactgatgactttgtggtcaatggcag
ctatgccgaacaaatgtacggaatgtgtgagtccctctgggaacccaacatggatccagaacacccgtttgaaaccatctc
cccagccatgctgaatgctgtggactggggtgcagggtcaggcatgggagtcatcatccacatcaccaagaaggacaaa
atcaccaccaggaca

Appendix II

MOOSE13873_ctgchr11q_1	A1 Pepsin
MOOSE13874_ctgchr11q_1	A1 Pepsin
MOOSE13895_ctg10008	C1 Papain
MOOSE13908_ctgChr_13ctg3	C12 Ubiquitin C-terminal hydrolase family 1
MOOSE13930_ctg15ctg27	C15 Pyroglutamyl-peptidase I
MOOSE13940_ctgChr_Xctg264	C19 Ubiquitin C-terminal hydrolase family 2
MOOSE13941_ctgChr_Xctg264	C19 Ubiquitin C-terminal hydrolase family 2
MOOSE13943_ctgChr_Xctg26	C19 Ubiquitin C-terminal hydrolase family 2
MOOSE13948_ctg8ctg6	C19 Ubiquitin C-terminal hydrolase family 2
MOOSE13952_ctgChr_1ctg34	C19 Ubiquitin C-terminal hydrolase family 2
MOOSE13954_ctg8ctg4	C19 Ubiquitin C-terminal hydrolase family 2
MOOSE13975_ctg21fin2	C19 Ubiquitin C-terminal hydrolase family 2
MOOSE13977_ctgchr11q_5	C19 Ubiquitin C-terminal hydrolase family 2
MOOSE13980_ctg17005	C19 Ubiquitin C-terminal hydrolase family 2
MOOSE13982_ctgCHR12_11	C19 Ubiquitin C-terminal hydrolase family 2
MOOSE14044_ctg2013	M12A Astacin
MOOSE14065_ctgCHR12_3	M12B Reprolysin
MOOSE14068_ctg10007	M12B Reprolysin
MOOSE14071_ctgC22	M12B Reprolysin
MOOSE14072_ctg5007	M12B Reprolysin
MOOSE14097_ctgChr_1ctg2	M13 Neprilysin
MOOSE14111_ctg2019	M14 Carboxipeptidase Clan
MOOSE14139_ctg2019	M24A Methionyl aminopeptidase, type 1
MOOSE14190_ctg_2	S1 Chymotrypsin / trypsin
MOOSE14210_ctg22fin4	S1 Chymotrypsin / trypsin
MOOSE14212_ctg_2	S1 Chymotrypsin / trypsin
MOOSE14225_ctgCHR3_11	S1 Chymotrypsin / trypsin
MOOSE14231_ctg4012	S1 Chymotrypsin / trypsin
MOOSE14237_ctg2019	S1 Chymotrypsin / trypsin
MOOSE14238_ctg4012	S1 Chymotrypsin / trypsin
MOOSE14241_ctgCHR12_3	S1 Chymotrypsin / trypsin
MOOSE14247_ctgCHR12_2	S1 Chymotrypsin / trypsin
MOOSE14251_ctg4012	S1 Chymotrypsin / trypsin
MOOSE14254_ctg4015	S1 Chymotrypsin / trypsin
MOOSE14260_ctgchr7_ctg18037	S1 Chymotrypsin / trypsin
MOOSE14265_ctg2014	S1 Chymotrypsin / trypsin
MOOSE14272_ctgC20	S1 Chymotrypsin / trypsin
MOOSE14278_ctgChr_1ctg109	S1 Chymotrypsin / trypsin
MOOSE14279_ctgChr_6ctg20	S1 Chymotrypsin / trypsin
MOOSE14295_ctg11ctg15	S49 Endopeptidase IV (sppA) (E.coli)
MOOSE14296_ctgC31	S49 Endopeptidase IV (sppA) (E.coli)
MOOSE14301_ctg_2	S8 Subtilases
MOOSE14331_ctg18ctg2	T1A Threonine Type Peptidases
MOOSE14335_ctg18ctg2	T1A Threonine Type Peptidases

MOOSE14337_ctgChr_9ctg1926	U22 Drosophila transposon 297 endopeptidase
MOOSE14342_ctgchr14_1	T1B Threonine Type Peptidases
MOOSE14348_ctg2019	T1B Threonine Type Peptidases

Appendix III

PARKINSON'S DISEASE

Locus 1 Marker:D1S231 Lod:5.11 CM RANGE of one LOD drop: 7

MOOSE13952 C19 Ubiquitin C-terminal hydrolase family 2
DISTANCE: 12.78 Mb

Locus 3 Marker:D1S2842 Lod:1.26 CM RANGE of one LOD drop: 51

MOOSE14278 S1 Chymotrypsin / trypsin
DISTANCE: -16.6 Mb

HYPERTENSION

Locus 4 Marker:D11S4102 Lod:1.5 CM RANGE of one LOD drop: 17

MOOSE14295 S49 Endopeptidase IV (sppA) (E.coli)
DISTANCE: -9.83 Mb

ANXIETY

Locus 1 Marker:D9S1690 Lod:4.38 CM RANGE of one LOD drop: 14

MOOSE14337 U22 Drosophila transposon 297 endopeptidase
DISTANCE: -9.03 Mb

COPD (CHRONIC OBSTRUCTIVE PULMONARY DISEASE)

Locus 4 Marker:D19S884 Lod:2.9 CM RANGE of one LOD drop: 20

MOOSE14301 S8 Subtilases
DISTANCE: -8.63 Mb
MOOSE14190 S1 Chymotrypsin / trypsin
DISTANCE: -7.29 Mb
MOOSE14212 S1 Chymotrypsin / trypsin
DISTANCE: -7.24 Mb

Locus 5 Marker:D21S1884 Lod:3.7 CM RANGE of one LOD drop: 12

MOOSE13975 C19 Ubiquitin C-terminal hydrolase family 2
DISTANCE: -5.37 Mb

ASTHMA

Locus 2 Marker:D3S1546 Lod:3.4 CM RANGE of one LOD drop: 17

MOOSE14225 S1 Chymotrypsin / trypsin
DISTANCE: 8.894 Mb

NIDDM (NON-INSULIN DEP. DIABETES)

Locus 3 Marker:D12S79 Lod:3.6 CM RANGE of one LOD drop: 12

MOOSE13982 C19 Ubiquitin C-terminal hydrolase family 2
DISTANCE: -6.67 Mb

OBESITY

Locus 3 Marker:D12S79 Lod:3.9 CM RANGE of one LOD drop: 10

MOOSE13982 C19 Ubiquitin C-terminal hydrolase family 2
DISTANCE: -6.67 Mb

Locus 4 Marker:D14S283 Lod:3.2 CM RANGE of one LOD drop: 12

MOOSE14342 T1B Threonine Type Peptidases
DISTANCE: 0.824 Mb

MIGRAINE (GENOME-WIDE SCAN ONLY)

Locus 1 Marker:D2S347 Lod:2.8 CM RANGE of one LOD drop: 45

MOOSE14265 S1 Chymotrypsin / trypsin
DISTANCE: 7.808 Mb

Locus 2 Marker:D2S2321 Lod:2.0 CM RANGE of one LOD drop: 25

MOOSE14111 M14 Carboxipeptidase Clan
DISTANCE: -0.46 Mb
MOOSE14348 T1B Threonine Type Peptidases
DISTANCE: 9.374 Mb

BIPOLAR (GENOME-WIDE SCAN ONLY)

Locus 1 Marker:D1S434 Lod:3.3 CM RANGE of one LOD drop: 25

MOOSE14097 M13 Neprilysin
DISTANCE: -12.7 Mb

OP (OSTEOPOROSIS)

Locus 2 Marker:D11S4102 Lod:2.95 CM RANGE of one LOD drop: 11

MOOSE14295 S49 Endopeptidase IV (sppA) (E.coli)
DISTANCE: -9.83 Mb

MYOPIA

Locus 1 Marker:D2S2215 Lod:1.34 CM RANGE of one LOD drop: 30

MOOSE14265 S1 Chymotrypsin / trypsin
DISTANCE: 1.613 Mb

ALZHEIMER'S DISEASE

Locus 3 Marker:D13S789 Lod:4.12 CM RANGE of one LOD drop: 4

MOOSE13908 C12 Ubiquitin C-terminal hydrolase family 1
DISTANCE: -2.63 Mb

PAOD (PERIPHERAL ARTERIAL OCCLUSIVE DISEASE)

Locus1 Marker:D1S2798 Lod:4.32 CM RANGE of one LOD drop: 3

MOOSE13952 C19 Ubiquitin C-terminal hydrolase family 2
DISTANCE: -8.16 Mb

Locus2 Marker:D1S2846 Lod:2.42 CM RANGE of one LOD drop: 8

MOOSE13952 C19 Ubiquitin C-terminal hydrolase family 2
DISTANCE: 1.451 Mb

AMD (AGE-RELATED MACULAR DEGENERATION)

Locus3 Marker:D16S3057 Lod:2.84 CM RANGE of one LOD drop: 14

MOOSE14272 S1 Chymotrypsin / trypsin
DISTANCE: 7.521 Mb

PROSTATE CANCER

Locus2 Marker:D21S1916 Lod:1.21 CM RANGE of one LOD drop: 38

MOOSE13975 C19 Ubiquitin C-terminal hydrolase family 2
DISTANCE: -11.7 Mb

MI (MYOCARDIAL INFARCTION)

Locus2 Marker:D7S2513 Lod:1.75 CM RANGE of one LOD drop: 22

MOOSE14260 S1 Chymotrypsin / trypsin
DISTANCE: 0.603 Mb

OSTEOARTHRITIS

Locus2 Marker:D4S2999 Lod:3.8 CM RANGE of one LOD drop: 11

MOOSE14254 S1 Chymotrypsin / trypsin
DISTANCE: -2.48 Mb

Locus4 Marker:D16S3255 Lod:3 CM RANGE of one LOD drop: 9

MOOSE14272 S1 Chymotrypsin / trypsin
DISTANCE: 10.31 Mb

IBD-INFLAMMATORY BOWEL DISEASE

Locus4 Marker:D16S3040 Lod:2.2 CM RANGE of one LOD drop: 16

MOOSE14071 M12B Reprolysin
DISTANCE: -2.31 Mb
MOOSE14296 S49 Endopeptidase IV (sppA) (E.coli)
DISTANCE: 12.52 Mb

LONGEVITY

Locus2 Marker:D11S987 Lod:2.07 CM RANGE of one LOD drop: 12

MOOSE13874 A1 Pepsin
DISTANCE: -9.92 Mb
MOOSE13873 A1 Pepsin
DISTANCE: -9.81 Mb

CLAIMS

What is claimed is:

1. An isolated nucleic acid molecule comprising a protease gene, wherein the protease gene has a nucleotide sequence selected from the group of nucleic acid sequences as shown in Appendix I, or the complements of the group of nucleic acid sequences as shown in Appendix I.

2. A nucleic acid encoding a polypeptide, wherein the polypeptide has an amino acid sequence selected from the group consisting of the group of amino acid sequences as shown in Appendix I.

3. An isolated nucleic acid molecule which hybridizes under high stringency conditions to a nucleotide sequence selected from the group of nucleic acid sequences as shown in Appendix I, or the complements of the group of nucleic acid sequences as shown in Appendix I.

4. An isolated nucleic molecule which hybridizes under high stringency conditions to a nucleotide sequence encoding an amino acid sequence selected from the group consisting of the group of amino acid sequences as shown in Appendix I.

5. A method for assaying for the presence of a first nucleic acid molecule in a sample, comprising contacting said sample with a second nucleic acid molecule, where the second nucleic acid molecule comprises a nucleotide sequence selected from the group of nucleic acid sequences as shown in Appendix I, and hybridizes to the first nucleic acid under high stringency conditions.

6. A vector comprising an isolated nucleic acid molecule selected from the group consisting of:

(a) the nucleic acid sequences as shown in Appendix I;

(b) the complement of one of the nucleic acid sequences are shown in Appendix I; or

(c) a nucleic acid encoding an amino acid molecule as shown in Appendix I;

where the nucleic acid molecule is operably linked to a regulatory sequence.

7. A recombinant host cell comprising the vector of Claim 6.

8. A method for producing a polypeptide encoded by an isolated nucleic acid molecule, comprising culturing the recombinant host cell of Claim 7 under conditions suitable for expression of the nucleic acid molecule.

9. An isolated polypeptide encoded by the nucleotide sequence of the group of nucleic acid sequences as shown in Appendix I, or the complements thereof.

10. The isolated polypeptide of Claim 9, wherein the polypeptide has an amino acid sequence selected from the group consisting of the group of amino acid sequences as shown in Appendix I.

11. An isolated polypeptide comprising an amino acid sequence, wherein the amino acid sequence is greater than about 95% identical to an amino acid sequence selected from the group consisting of the group of amino acid sequences as shown in Appendix I.

12. A fusion protein comprising an isolated polypeptide of Claim 2.

13. A fusion protein comprising an isolated polypeptide of Claim 11.

14. An antibody, or an antigen-binding fragment thereof, which selectively binds to a polypeptide of Claim 2, or to a fragment or variant of said amino acid sequence.

15. An antibody, or an antigen-binding fragment thereof, which selectively binds to a polypeptide of Claim 11, or to a fragment or variant of said amino acid sequence.

16. A method of assaying for the presence of a polypeptide encoded by an isolated nucleic acid molecule according to Claim 1 in a sample, the method comprising contacting the sample with an antibody which specifically binds to the encoded polypeptide.

17. A method of identifying an agent which alters the activity of a protease, the method comprising:

(a) contacting a polypeptide of Claim 9, or a derivative or fragment thereof, with an agent to be tested;

(b) assessing the level of activity of the polypeptide or derivative or fragment thereof; and

(c) comparing the level of activity with a level of activity of the polypeptide or active derivative or fragment thereof in the absence of the agent;

wherein if the level of activity of the polypeptide or derivative or fragment thereof in the presence of the agent differs, by an amount that is statistically significant, from the level in the absence of the agent, then the agent is an agent that alters activity of a protease.

18. An agent which alters the activity of a protease, identifiable according to the method of Claim 17.

19. The agent of Claim 18, where the agent is selected from the group consisting of: a protease gene binding agent; a receptor; a peptidomimetic; a fusion protein; a prodrug; an antibody; and a ribozyme.

20. A method of altering activity of a polypeptide encoded by a protease gene, comprising contacting the polypeptide with an agent of Claim 19.

21. A method of identifying an agent which alters interaction of the polypeptide of Claim 9 with a protease gene binding agent, comprising:

a) contacting the polypeptide or a derivative or fragment thereof, and the binding agent, with an agent to be tested;

b) assessing the interaction of the polypeptide or derivative or fragment thereof with the binding agent; and

c) comparing the level of interaction with a level of interaction of the polypeptide or derivative or fragment thereof with the binding agent in the absence of the agent,

wherein if the level of interaction of the polypeptide or derivative or fragment thereof in the presence of the agent differs by an amount that is statistically significant, from the level of interaction in the absence of the agent, then the agent is an agent that alters interaction of the polypeptide with the binding agent.

22. An agent which alters interaction of a protease gene polypeptide with a protease gene binding agent, identifiable according to the method of Claim 21.

23. An agent which alters interaction of a protease gene polypeptide with a protease gene binding agent, selected from the group consisting of: a second protease gene binding agent; a receptor; a peptidomimetic; a fusion protein; a prodrug; an antibody; and a ribozyme.

24. A method of altering interaction of a protease gene polypeptide with a protease gene binding agent, comprising contacting the protease gene polypeptide and/or the protease gene binding agent with an agent of Claim 23.

25. A method of identifying an agent which alters expression of a protease gene, comprising the steps of:

a) contacting a solution containing a nucleic acid comprising the promoter region of the protease gene operably linked to a reporter gene with an agent to be tested;

b) assessing the level of expression of the reporter gene; and

c) comparing the level of expression with a level of expression of the reporter gene in the absence of the agent,

wherein if the level of expression of the reporter gene in the presence of the agent differs, by an amount that is statistically significant, from the level of expression in the absence of the agent, then the agent is an agent that alters expression of the protease gene.

26. An agent which alters expression of the protease gene, identifiable according to the method of Claim 25.

27. A method of identifying an agent which alters expression of a protease gene, comprising the steps of:

a) contacting a solution containing a nucleic acid of Claim 1 or a derivative or fragment thereof with an agent to be tested;

b) assessing expression of the nucleic acid, derivative or fragment; and

c) comparing expression with expression of the nucleic acid, derivative or fragment in the absence of the agent,

wherein if expression of the nucleotide, derivative or fragment in the presence of the agent differs, by an amount that is statistically significant,

from the expression in the absence of the agent, then the agent is an agent that alters expression of the protease gene.

28. The method of Claim 27, wherein the expression of the nucleotide, derivative or fragment in the presence of the agent comprises expression of one or more splicing variant(s) that differ in kind or in quantity from the expression of one or more splicing variant(s) the absence of the agent.

29. An agent which alters expression of a protease gene, identifiable according to the method of Claim 27.

30. An agent which alters expression of a protease gene, selected from the group consisting of: antisense nucleic acid to a protease gene; a protease gene polypeptide; a protease gene receptor; a protease gene binding agent; a peptidomimetic; a fusion protein; a prodrug thereof; an antibody; and a ribozyme.

31. A method of altering expression of a protease gene, comprising contacting a cell containing a protease gene with an agent of Claim 30.

32. A method of identifying a polypeptide which interacts with a protease gene polypeptide, comprising employing a two yeast hybrid system using a first vector which comprises a nucleic acid encoding a DNA binding domain and a protease gene polypeptide, splicing variant, or a fragment or derivative thereof, and a second vector which comprises a nucleic acid encoding a transcription activation domain and a nucleic acid encoding a test polypeptide, wherein if transcriptional activation occurs in the two yeast hybrid system, the test polypeptide is a polypeptide which interacts with a protease polypeptide.

33. A protease gene therapeutic agent selected from the group consisting of: a protease gene or fragment or derivative thereof; a polypeptide encoded by a protease gene; a receptor; a protease gene binding agent; a peptidomimetic; a fusion protein; a prodrug; an antibody; an agent that alters protease gene expression; an agent that alters activity of a polypeptide encoded by a protease gene; an agent that alters posttranscriptional processing of a polypeptide encoded by a protease gene; an agent that alters interaction of a protease gene with a protease gene binding agent; an agent that alters transcription of splicing variants encoded by a protease gene; and a ribozyme.

34. A pharmaceutical composition comprising a protease gene therapeutic agent of Claim 33.

35. The pharmaceutical composition of Claim 34, wherein the protease gene therapeutic agent is an isolated nucleic acid molecule comprising a protease gene or fragment or derivative thereof.

36. The pharmaceutical composition of Claim 34, wherein the protease gene therapeutic agent is a polypeptide encoded by the protease gene.

37. A method of treating a disease or condition associated with a protease in an individual, comprising administering a protease gene therapeutic agent to the individual, in a therapeutically effective amount.

38. The method of Claim 37, wherein the protease gene therapeutic agent is a protease gene agonist.

39. The method of Claim 38 wherein the protease gene therapeutic agent is a protease gene antagonist.

40. A transgenic animal comprising a nucleic acid selected from the group consisting of: an exogenous protease gene and a nucleic acid encoding a protease gene polypeptide.

41. A method for assaying a sample for the presence of a protease gene nucleic acid, comprising:

a) contacting said sample with a nucleic acid comprising a contiguous nucleotide sequence which is at least partially complementary to a part of the sequence of said protease gene nucleic acid under conditions appropriate for hybridization, and

b) assessing whether hybridization has occurred between a protease gene nucleic acid and said nucleic acid comprising a contiguous nucleotide sequence which is at least partially complementary to a part of the sequence of said protease gene nucleic acid;

where if hybridization has occurred, a protease gene is present in the nucleic acid.

42. The method of Claim 41, wherein said nucleic acid comprising a contiguous nucleotide sequence is completely complementary to a part of the sequence of said protease gene nucleic acid.

43. The method of Claim 41, comprising amplification of at least part of said protease gene nucleic acid.

44. The method of Claim 41, wherein said contiguous nucleotide sequence is 100 or fewer nucleotides in length and is either: a) at least 80% identical to a contiguous sequence of nucleotides in one of the nucleic acid sequences as shown in Appendix I; b) at least 80% identical to the complement of a contiguous sequence of nucleotides in one of the nucleic acid sequences as shown in Appendix I; or c) capable of selectively hybridizing to said protease gene nucleic acid.

45. A reagent for assaying a sample for the presence of a protease gene nucleic acid, said reagent comprising a nucleic acid comprising a contiguous nucleotide sequence which is at least partially complementary to a part of the nucleotide sequence of said protease gene nucleic acid.

46. The reagent of Claim 45, wherein the nucleic acid comprises a contiguous nucleotide sequence which is completely complementary to a part of the nucleotide sequence of said protease gene nucleic acid.

47. A reagent kit for assaying a sample for the presence of a protease gene nucleic acid, comprising in separate containers:

a) one or more labeled nucleic acids comprising a contiguous nucleotide sequence which is at least partially complementary to a part of the nucleotide sequence of said protease gene nucleic acid, and

b) reagents for detection of said label.

48. The reagent kit of Claim 47, wherein the labeled nucleic acid comprises a contiguous nucleotide sequences which is completely complementary to a part of the nucleotide sequence of said protease gene nucleic acid.

49. A reagent kit for assaying a sample for the presence of a protease gene nucleic acid, comprising one or more nucleic acids comprising a contiguous nucleotide sequence which is at least partially complementary to a part of the nucleotide sequence of said protease gene nucleic acid, and which is capable of acting as a primer for said protease gene nucleic acid when maintained under conditions for primer extension.

50. The use of a nucleic acid which is 100 or fewer nucleotides in length and which is either: a) at least 80% identical to a contiguous sequence of nucleotides in one of the nucleic acid sequences as shown in Appendix I; b) at least 80% identical to the complement of a contiguous sequence of

nucleotides in one of the nucleic acid sequences as shown in Appendix I; or c) capable of selectively hybridizing to said protease gene nucleic acid, for assaying a sample for the presence of a protease gene nucleic acid.

51. The use of a first nucleic acid which is 100 or fewer nucleotides in length and which is either:

a) at least 80% identical to a contiguous sequence of nucleotides in one of the nucleic acid sequences as shown in Appendix I;

b) at least 80% identical to the complement of a contiguous sequence of nucleotides in one of the nucleic acid sequences as shown in Appendix I; or

c) capable of selectively hybridizing to said protease gene nucleic acid;

for assaying a sample for the presence of a protease gene nucleic acid that has at least one nucleotide difference from the first nucleic acid.

52. The use of a nucleic acid which is 100 or fewer nucleotides in length and which is either:

a) at least 80% identical to a contiguous sequence of nucleotides in one of the nucleic acid sequences as shown in Appendix I;

b) at least 80% identical to the complement of a contiguous sequence of nucleotides in one of the nucleic acid sequences as shown in Appendix I; or

c) capable of selectively hybridizing to said protease gene nucleic acid;

for diagnosing a susceptibility to a disease or condition associated with a protease.

SEQUENCE LISTING

<110> deCODE genetics ehf.
Martinez, Roger Alfonso Moraga
Sigurdsson, Gunnar Thor

<120> NUCLEIC ACIDS ENCODING PROTEASES

<130> 2345.2035002

<150> 60/332,633
<151> 2001-11-06

<160> 94

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 1083
<212> DNA
<213> Homo sapiens

<400> 1
ctggtagatg aacagcccct ggagaactac ctggatatgg agtacttcgg cactatcggc 60
atcggaactc ctgcccagga tttcaccgtc gtctttgaca ccggctcctc caacctgtgg 120
gtgccctcag tctactgctc cagtcttgcc tgcaccaacc acaaccgctt caaccctgag 180
gattcttcca cctaccagtc caccagcgag acagtctcca tcacctacgg caccggcagc 240
atgacaggca tcctcggata cgacactgtc caggttggag gcatctctga caccaatcag 300
atcttcggcc tgagcgagac ggaacctggc tccttcctgt attatgctcc cttcgatggc 360
atcctggggc tggcctaccc cagcatttcc tcctccgggg ccacacccgt ctttgacaac 420
atctggaacc agggcctggt ttctcaggac ctcttctctg tctacctcag cgccgatgac 480
aagagtggca gcgtggtgat ctttggtggc attgactctt cttactacac tggaagtctg 540
aactgggtgc ctgttaccgt cgagggttac tggcagatca ccgtggacag catcaccatg 600
aacggagaga ccatcgcctg tgctgagggc tgccaggcca ttgttgacac cggcacctct 660
ctgctgaccg gcccaaccag ccccattgcc aacatccaga gcgacatcgg agccagcgag 720
aactcagatg gcgacgtgag tccagccccg actgccctgt tctacactca agtagtgggt 780
gtgccaggca gaagcgacga aaacccttct aacttttctc accctcactc tttccagatg 840
gtggtcagct gctcagccat cagcagcctg cccgacatcg tcttcaccat caatggagtc 900
cagtaccccg tgccacccag tgcctacatc ctgcagagcg aggggagctg catcagtggc 960
ttccagggca tgaacgtccc caccgaatct ggagagcttt ggatcctggg tgatgtcttc 1020
atccgccagt actttaccgt cttcgacagg gcaaacaacc aggtcggcct ggcccctgtg 1080
gct 1083

<210> 2
<211> 1083
<212> DNA
<213> Homo sapiens

<400> 2
ctggtagatg aacagcccct ggagaactac ctggatatgg agtacttcgg cactatcggc 60
atcggaactc ctgcccagga tttcactgtc ctctttgaca ccggctcctc caacctgtgg 120
gtgccctcag tctactgctc cagtcttgcc tgcaccaacc acaaccgctt caaccctgag 180
gattcttcca cctaccagtc caccagcgag acagtctcca tcacctacgg caccggcagc 240
atgacaggca tcctcggata cgacactgtc caggttggag gcatctctga caccaatcag 300
atcttcggcc tgagcgagac ggaacctggc tccttcctgt attatgctcc cttcgatggc 360
atcctggggc tggcctaccc cagcatttcc tcctccgggg ccacacccgt ctttgacaac 420
atctggaacc agggcctggt ttctcaggac ctcttctctg tctacctcag cgccgatgac 480
aagagtggca gcgtggtgat ctttggtggc attgactctt cttactacac tggaagtctg 540
aactgggtgc ctgttaccgt cgagggttac tggcagatca ccgtggacag catcaccatg 600
aacggagaga ccatcgcctg tgctgagggc tgccaggcca ttgttgacac cggcacctct 660

ctgctgaccg gcccaaccag ccccattgcc aacatccaga gcgacatcgg agccagcgag 720
aactcagatg gcgacgtgag tccagccccg actgccctgt tctacactca agtagtgggt 780
gtgccaggca gaagcgacga aaacccttct aacttttctc accctcactc tttccagatg 840
gtggtcagct gctcagccat cagcagcctg cccgacatcg tcttcaccat caatggagtc 900
cagtaccccg tgccacccag tgcctacatc ctgcagagcg aggggagctg catcagtggc 960
ttccagggca tgaacgtccc caccgaatct ggagagcttt ggatcctggg tgatgtcttc 1020
atccgccagt actttaccgt cttcgacagg gcaaacaacc aggtcggcct ggcccctgtg 1080
gct 1083

<210> 3
<211> 915
<212> DNA
<213> Homo sapiens

<400> 3
gacaaggcat ggttgaaaag agggaacaaa cagttcaatg aaggaaagga atctgataga 60
tgtctgatat ttaaatgtaa aaataaggac gtgaagatga ttgagcagca caatcaggaa 120
tacagccaag ggaaacacag cttcacaatg gccatgaacg cctttggaga catgaccaat 180
gaagaattca ggcaggtgat gaatggtttt caataccaga agcacaggaa ggggaaacag 240
ttccaggaac gcctgcttct tgagatcccc acatctgtgg actggagaga gaaaggctac 300
atgactcctg tgaaggatca gggtcagtgt ggctcttgtt gggcttttag tgcaactggt 360
gctctggaag ggcagatgtt ctggaaaaca ggcaaactta tctcactgaa tgagcagaat 420
ctggtagact gctctgggcc tcaaggcaat gagggctgca atggtgactt catggataat 480
cccttccggt atgttcagga gaacggaggc ctggactctg aggcatccta tccatatgaa 540
ggaaaggtta aaacctgtag gtacaatccc aagtattctg ctgctaatga cactggtttt 600
gtggacatcc cttcacggga gaaggacctg gcgaaggcag tggcaactgt ggggcccatc 660
tctgttgctg ttggtgcaag ccatgtcttc ttccagttct ataaaaaagg aatttatttt 720
gagccacgct gtgaccctga aggcctggat catgctatgc tggtggttgg ctacagctat 780
gaaggagcag actcagataa caataaatat tggctggtga agaacaggga atccttttt 840
gcaatcagta tttacatacc acaagcccct aatcacattt ctgaaatccc agaagtcttt 900
ttccccctta ggcag 915

<210> 4
<211> 570
<212> DNA
<213> Homo sapiens

<400> 4
tcaaatgtga agccaagaaa gaaaaatcca aaagggttca ctgttaatgt cctggcaaac 60
attggtgtca taaatcactg gaaattaccc aaacttactc acaaagggga aaaaaatagt 120
ggaaaagagg gggaaaaacc acacaaaaca aacagtgcat ccaacagctg tgggacaatt 180
tcaaatttag gaagcattga aaatggaaca aatgaaggta aatataaaat aaattttatt 240
cttaaagtca ttttaaaaaa cgcaaacact gttagaactg aaaagagaaa tagacaaatt 300
cacaattata actgtttaca ctgcactgtt gaaacatcag attatcaagt tcaagcacca 360
agtatagatg agaaagtaga tcttcatttt attgcattag ttcatgtaga tgggcatctc 420
tatgaattag atgggcggaa gccatttcca attaaccatg gtgaaactag tgatgaaact 480
ttattagagg atgccataga agtttgcaag aagtttatgg agcgcgaccc tgatgaacta 540
agatttaatg cgattgctct ttctgcagca 570

<210> 5
<211> 612
<212> DNA
<213> Homo sapiens

<400> 5
gcatgtttgg tttcaggttt tgggcccttc cggcagcact tggtgaattc cagctgggaa 60
gcagtgaagg agctctccaa gctgggcctg gggaatgaaa cagtggtgca gctgcggact 120
ctggagctgc ctgtagatta cagggaggct aagcggaggg tcaccggaat ctgggaagat 180
catcagccgc aactcgtcgt gcatgtgggc atggacaccg ccgccaaggc gatcattctg 240
gaacagtctg gcaagaacca aggctaccgg gacgccgaca tccgcagctt ctggcccgag 300
ggcggcgtgt gcctacctgg cagcccagac gtgctggagt caggggtctg catgaaggca 360
gtctgcaagc gcgtagctgt ggagggtgtc gacgtgatct tttcccgaga tgcaggcaga 420

tacgtctgtg attataccta ttacctgtct ctgcatcatg gaaagggctg cgcggcactc 480
atccatgtcc ctccactatc gcgcgggctc ccggccagcc tgctgggaag agccttgaga 540
gtcatcatcc aggaaatgct ggaagaggca ggggagaaac aaaaagaggt gacggcttct 600
ggcacatccc ac 612

<210> 6
<211> 1074
<212> DNA
<213> Homo sapiens

<400> 6
agaaaaaagt cagtctatac tgtaggcctg agagggctaa tcaatcttgg gaacacttgt 60
tttatgaatt gtattgtcca ggcacttacc catattcctc tactgaaaga tttcttcctc 120
tctgacaagc acaaatgtat aatgacaagc cccagcttgt gtctggtctg tgaaatgtct 180
tcgcttttc atgctatgta ctctgggagc cgaactcctc acattcccta taagttactg 240
catctgatat ggatccatgc agaacattta gcagggtaca ggcagcagga tgcccatgag 300
ttccttattg caatattaga cgtgctacat agacacagca aagatgatag tggtgggcag 360
gaggccaata accccaactg ctgtaactgc atcatagacc aaatctttac aggtggcctg 420
caatcagatg tcacatgtca agcctgccat agtgtttcta ccaccataga cccatgctgg 480
gacatcagtt tggacttgcc tggctcttgt gccacattcg attcccagaa cccagagagg 540
gctgacagca cagtgagcag ggatgaccac ataccaggaa tcccctcact tacagactgt 600
ctacagtggt ttacaaggcc agagcaccta ggaagcagtg ccaaaatcaa atgcaatagt 660
tgccaaagct accaggagtc tactaaacag ctcacaatga aaaaattacc cattgtggct 720
tgttttcatc tcaagcggtt tgagcatgta ggcaaacaga ggcgaaagat taataccttt 780
atctcctttc ccttggagct ggacatgact ccgtttttgg cctctactaa agagagcaga 840
atgaaagaag gccagccacc aacagattgt gtgcccaatg agaataagta ttccttgttt 900
gcagtgatta atcaccatgg aactttggaa agtggccact ataccagctt catccggcaa 960
caaaaggacc agtggttcag ctgtgatgat gccatcatca ccaaggctac cattgaggac 1020
ttactctaca gtgaagggta tttactgttc tatcacaaac agggtctaga gaaa 1074

<210> 7
<211> 1074
<212> DNA
<213> Homo sapiens

<400> 7
agaaaaaagt cagtctatac tgtaggcctg agagggctaa tcaatcttgg gaacacttgt 60
tttatgaatt gtattgtcca ggcacttacc catattcctc tactgaaaga tttcttcctc 120
tctgacaagc acaaatgtat aatgacaagc cccagcttgt gtctggtctg tgaaatgtct 180
tcgcttttc atgctatgta ctctgggagc cgaactcctc acattcccta taagttactg 240
catctgatat ggatccatgc agaacattta gcagggtaca ggcagcagga tgcccatgag 300
ttccttattg caatattaga cgtgctacat agacacagca aagatgatag tggtgggcag 360
gaggccaata accccaactg ctgtaactgc atcatagacc aaatctttac aggtggcctg 420
caatcagatg tcacatgtca agcctgccat agtgtttcta ccaccataga cccatgctgg 480
gacatcagtt tggacttgcc tggctcttgt gccacattcg attcccagaa cccagagagg 540
gctgacagca cagtgagcag ggatgaccac ataccaggaa tcccctcact tacagactgt 600
ctacagtggt ttacaaggcc agagcaccta ggaagcagtg ccaaaatcaa atgcaatagt 660
tgccaaagct accaggagtc tactaaacag ctcacaatga aaaaattacc cattgtggct 720
tgttttcatc tcaagcggtt tgagcatgta ggcaaacaga ggcgaaagat taataccttt 780
atctcctttc ccttggagct ggacatgact ccgtttttgg cctctactaa agagagcaga 840
atgaaagaag gccagccacc aacagattgt gtgcccaatg agaataagta ttccttgttt 900
gcagtgatta atcaccatgg aactttggaa agtggccact ataccagctt catccggcaa 960
caaaaggacc agtggttcag ctgtgatgat gccatcatca ccaaggctac cattgaggac 1020
ttactctaca gtgaagggta tttactgttc tatcacaaac agggtctaga gaaa 1074

<210> 8
<211> 1071
<212> DNA
<213> Homo sapiens

<400> 8
agaatcacct ccagctttac gatcggttta agaggactca tcaatcttgg caacacgtgc 60

tttatgaact gcattgtcca ggccctcacc cacacgccga tactgagaga tttctttctc 120
tctgacaggc accgatgtga gatgccgagt cccgagttgt gtctggtctg tgagatgtcg 180
tcgctgtttc gggagttgta ttctggaaac ccgtctcctc atgtgcccta taagttactg 240
cacctggtgt ggatacatgc ccgccattta gcagggtaca ggcaacagga tgcccacgag 300
ttcctcattg cagcgttaga tgtcctgcac aggcactgca aaggtgatga tgtcgggaag 360
gcggccaaca atcccaacca ctgtaactgc atcatagacc aaatcttcac aggtggcctg 420
cagtctgatg tcacctgtca agcctgccat ggcgtctcca ccacgataga cccatgctgg 480
gacattagtt tggacttgcc tggctcttgc acctccttct ggcccatgag cccagggagg 540
gagagcagtg tgaacgggga aagccacata ccaggaatca ccaccctcac ggactgcttg 600
cggaggttta cgaggccaga gcacttagga agcagtgcca aaatcaaatg tggtagttgc 660
caaagctacc aggaatctac caaacagctc acaatgaata aattacctgt cgttgcctgt 720
tttcatttca aacggtttga acattcagcg aaacagaggc gcaagatcac tacatacatt 780
tcctttcctc tggagctgga tatgacgccg tttatggcct caagtaaaga gagcagaatg 840
aatggacaat tgcagctgcc aaccaatagt ggaaacaacg aaaataagta ttccttgttt 900
gctgtggtta atcaccaagg aaccttggag agtggccact ataccagctt catccggcac 960
cacaaggacc agtggttcaa gtgtgatgat gccgtcatca ctaaggccag tattaaggac 1020
gtactggaca gtgaagggta tttactgttc tatcacaaac aggtgctaga a 1071

<210> 9
<211> 924
<212> DNA
<213> Homo sapiens

<400> 9
ctgagtagca ggagacctgc tgcggtgggg gctgggctcc agaatatggg aaatacctgc 60
tacgagaacg cttccctgca gtgcctgaca tacacaccgc cccttgccaa ctacatgctg 120
tcccgggagc actctcaaac atgtcagcgt cccaagtgct gcatgctctg tactatgcaa 180
gctcacatca catgggccct ccacagtcct ggtcatgtca tccagccctc acaggcattg 240
gctgctggct tccatagagg caagcaggaa gatgcccatg aatttctcat gttcactgtg 300
gatgccatga aaaaggcatg ccttcccggc cacaagcagg tagatcatca ctctaaggac 360
accaccctca tccaccaaat atttggaggc tgctggagat ctcaaatcaa gtgtctccac 420
tgccacggga tttcagacac ttttgaccct tacctggaca tcgccctgga tatccaggca 480
gctcagagtg tcaagcaagc tttggaacag ttggtgaagc ccgaagaact caatggagag 540
aatgcctatc attgcggtct ttgtctccag agggcgccag cctccaagac gttaacttta 600
cacacttctg ccaaggtcct catccttgtc ttgaagagat tctccgatgt cacaggcaac 660
aaacttgcca agaatgtgca atatcctgag tgccttgaca tgcagccata catgtctcag 720
cagaacacag gacctcttgt ctatgtcctc tatgctgtgc tggtccacgc tgggtggagt 780
tgtcacgatg gacattactt ctcttatgtc aaagctcaag aaggccagtg gtataaaatg 840
gatgatgcca aggtcactgc ctgtagcatc acttctgtcc tgagtcaaca ggcctatgtc 900
ctcttttaca tccagaagag tgaa 924

<210> 10
<211> 1011
<212> DNA
<213> Homo sapiens

<400> 10
tgtgagaaga gagaaaactt gttaccattt gtgggactga ataatctcgg caatacttgc 60
tatcttaata gtatacttca ggtattatat ttttgtcccg gctttatgta ctgtatcttc 120
aaaactagaa tagatgaaat ggaaattttt atttataggg aactcaaccc tatgtatgaa 180
ggatatctac agcatgatgc acaggaagta ttacaatgta ttttgggaaa cattcaagaa 240
acatgccaac tcctaaaaaa agaagaagta aaaaatgtgg cagaattacc tactaaggta 300
gaagaaatac ctcatccgaa agaggaaatg aatggtgaag aacaaattgg ttttgagcta 360
gtggagaaat tatttcaagg tcagctggta ttaaggacgc gttgcttgga atgtgaaagt 420
ttaacagaaa gaagagaaga ttttcaagac atcagtgtgc cagtacaaga agatgagctt 480
tccaaagtag aggagagttc tgaaaaaatg aagaccctga gatgggcaat ttcacaattt 540
gcttcagtag aaaggattgt aggagaagat aaatatttct gtgaaaactg ccatcattat 600
actgaagctg aacgaagtct tttgtttgac aaaatgcctg aagttataac tattcatttg 660
aagtgctttg ctgctagtgg aaacctttte ttttttcct cccaaaggtt tgattgttat 720
ggtggtggac tttccaagat caacactcct ttattgacac ctcttaaatt gtcactagaa 780
gaatggagca caaagccaac taacgacagc tatggattat ttgcggttgt gatgcatagt 840
ggcattacaa ttagtagtgg gcattacact gcttctgtta aagagtatga ggggaagtgg 900

ttgctttttg atgattctga agtcaaagtt actgaagaga aggactttct gaattctctt 960
tccccttcta catctcctac ttctactcct tacttgctat tttataagaa a 1011

<210> 11
<211> 924
<212> DNA
<213> Homo sapiens

<400> 11
ctgagtagca ggagacctgc tgcggtgggg gctgggctcc agaatatggg aaatacctgc 60
tacgagaacg cttccctgca gtgcctgaca tacacactgc cccttgccaa ctacatgctg 120
tcccgggagc actctcaaac atgtcagcgt cccaagtgct gcatgctctg tactatgcaa 180
gctcacatca catgggccct ccacagtcct ggccatgtca tccagccctc acaggcattg 240
gctgctggct tccatagagg caagcaggaa gatgtccatg aatttctcat gttcactgtg 300
gatgccatga aaaaggcatg ccttcccggc cacaagcagg tagatcatca ctgcaaggac 360
accaccctca tccaccaaat atttggaggc tgctggagat ctcaaatcaa gtgtctccac 420
tgccacggga tttcagacac ttttgaccct tacctggaca tcgccctgga tatccaggca 480
gctcagagtg tcaagcaagc tttggaacag ttggtgaagc ccgaagaact caatggagag 540
aatgcctatc attgcggtct ttgtctccag agggcgccgg cctccaacac gttaacttta 600
cacacttctg ccaaggtcct catccttgtc ttgaagagat tctccgatgt cgcaggcaac 660
aaacttgcca agaatgtgca atatcctgag tgccttgaca tgcagccata catgtctcag 720
cagaacacag gacctcttgt ctatgtcctc tatgctgtgc tggtccacgc tgggtggagt 780
tgtcacgacg gacattactt ctcctatgtc aaagctcaag aagtccagtg gtataaaatg 840
gatgatgccg aggtcactgt ctgtagcatc atttctgtcc tgagtcaaca ggcctatgtc 900
ctcttttaca tccagaagag tgaa 924

<210> 12
<211> 696
<212> DNA
<213> Homo sapiens

<400> 12
tatgatagaa aaagacagga caaagctccc gttgggctaa agaatgttgg caatacttgt 60
tggtttagtg ctgttattca gtcattattt aatcttttgg aatttagaag attagttctg 120
aattacaagc ctccatcaaa tgctcaagat ttaccccgaa accaaaaggc ttttttcttt 180
tcacagcaag atgtgagtga gtttacacac aaattattag attggttaga agatgccttc 240
caaatgaaag ctgaagagga gacggttggc aaggatgtgg agaaattgaa acccttgtgc 300
agtgttggtg aggatatgaa atggtacagc cactgtggaa aacatttttg ttattgtttt 360
atttcctttc agcattggtt tactgaatta ccacctgtgt taacatttga attgtcaaga 420
tttgaattta atcaggcatt gggaagacca gaaaaaattc acaacaaatt agaatttccc 480
caagttcctt atcgattaca tgccgtttta gttcacgaag gccaagctaa tgctgggcac 540
tactgggcat atatttttga tcatcgtgaa agcagatgga tgaagtacaa tgatattgct 600
gtgacaaaat catcatggga agagctagtg agggactctt ttggtggtta tagaaatgcc 660
agtgcatact gtttaatgta cataaatgat aaggca 696

<210> 13
<211> 684
<212> DNA
<213> Homo sapiens

<400> 13
aatgactgga ggagagttga tggttggcca gttgggctga aaaatgttgg caatacatgt 60
tggtttagtg ctgttattca gtctctcttt caattgcctg aatttcgaag acttgttctc 120
agttatagtc tgccacaaaa tgtacttgaa aattgtcgaa gtcatacaga acagcagcaa 180
gatgtgagtg aattcacaca caagctcctg gattggctag aggacgcatt ccagctagct 240
gttaatacct tcggccagta tcctcttcag gtaaacggtt atcgcaactt agacgagtgt 300
ttggaagggg ccatggtgga gggtgatgtt gagcttcttc cctccgatca ctcggtgaag 360
tatggacaag agcgttggtt tacaaagcta cctccagtgt tgacctttga actctcaaga 420
tttgagttta atcactcctg gggcagagac aagaaggatt caaaggcctt gcacacagtg 480
ccttatcgct tgcatgcagt tcttgttcat gaaggacaag caaatgctgg acactattgg 540
gcctatatct ataatcaacc ccgacagagc tggctcaagt acaatgacat ctctgttact 600
gaatcttcct gggaagaagt tgaaagagat tcctatggag gcctgagaaa tgttagtgct 660

tactgtctga tgtacattaa tgac 684

<210> 14
<211> 768
<212> DNA
<213> Homo sapiens

<400> 14
tctggaagct ctccttccag ttcatggcct tctggcctcc ggtcgagctg tcccattttt 60
caatgtttgt ttatgttgca tttgctttcg aggtctcagt cattcctctg gccaagagtg 120
aggatgagga ggcagcatgg tgctctggaa tttcatagag tcctgttcgg gagcctccag 180
gaggagcgag cgcaggatgc cgacagtgtg tggcagcagc agcaggcgca tcagcagcac 240
agctgtacct tggatgaatg ttttcagttc tacaccaagg aggagcagct ggcccaggat 300
gacgcctgga agtgtcctca ctgccaagtc ctgcagcagg ggatggtgaa gctgagtttg 360
tggacgctgc ctgacatcct catcatccac ctcaaaaggt tctgccaggt gggcgagaga 420
agaaacaagc tctccacgct ggtgaagttt ccgctctctg gactcaacat ggctccccat 480
gtggcccaga gaagcaccag ccctgaggca ggactgggcc cctggccttc ctggaagcag 540
ccggactgcc tgcccaccag ttacccgctg gacttcctgt acgacctgta tgccgtctgc 600
aaccaccatg gcaacctgca aggtgggcat tacacagcct actgccggaa ctctctggat 660
ggccagtggt acagttatga tgacagcacg gtggaaccgc ttcgagaaga tgaggtcaac 720
accagagggg cttatatcct gttctatcag aagcggaaca gcatccct 768

<210> 15
<211> 921
<212> DNA
<213> Homo sapiens

<400> 15
cttccccctg ctttttttt agggcttgtg cctggccttg ttaatttagg gaacacctgc 60
ttcatgaact ccctgctaca aggcctgtct gcctgtcctg ctttcatcag ccttgcctta 120
tttatctttg aatctctact tccactctac agctgcagtt ttatagccca ggagggaatt 180
catttatata ggcaacaaga tgctcacgaa ttattccatg tcattacctc gtcattggaa 240
gatgagcgag accgccagcc tcggtctcaa catccttttc atggaagact cactagtaat 300
atggtctgca aacactgtga acaccagagt cctgttcgat ttgatacctt tgatagcctt 360
tcactaagta ttccagccgc cacatggggt cacccattga ccctggacca ctgccttcac 420
cacttcatct catcagaatc agtgcgggat gttgtgtgtg acaactgtac aaagaggacc 480
acttttgtta aacagttaaa actagggaag gtgagcccac actacacacc ctgttggctt 540
tgttttgagg actccgtgta tcctgcccct gaaacaactc ggttctcccg atttctcttc 600
cacccgcagc tccctcagtg tctctgcatc cacctacagc ggctgagctg gtccagccac 660
ggcacgcctc tgaagcggca tgagcacgtg cactcctcca catacctctt ccggctgatg 720
gcagttgtcg tccaccatgg agacatgcac tctggacact ttgtcactta ccgacggtcc 780
ccaccttctg ccaggaaccc tctctcaact agcaatcagt ggctgtgggt ctccgatgac 840
actgtccgca aggccagcct gcaggaggtc ctgtcctcca gcgcctacct gctgttctac 900
gagcgcgtcc tttccaggat g 921

<210> 16
<211> 948
<212> DNA
<213> Homo sapiens

<400> 16
atgctctcct cccccgattt ctacccagct taccctctg caatgcaatc tctaagcctt 60
ggctccttgg ccagggcctt agaactgatg acccagtact ttaacaattg gaactgggtc 120
tacgacaaca taatagatca gaatgaaagc aaattgagta agtcaaggag agaagagata 180
gagagagaca gagagagaaa agagagaaga gagggagaca gggaaaaaaa gagacagaat 240
ggtgtcgtgg aggtcccctt cctgctctcc agcaagtacg atgagcccag ccgccaggtc 300
atcctggagg ctcttgcgga gtttgaacgt tccacgtgca tcaggtttgt cacctatcag 360
gaccagagag acttcatttc catcatcccc atgtatgggt gcttctcgag tgtggggcgc 420
agtggaggga tgcaggtggt ctccctggcg cccacgtgtc tccagaaggg ccggggcatt 480
gtccttcatg agctcatgca tgtgctgggc ttctggcacg agcacacgcg ggccgaccgg 540
gaccgctata tccgtgtcaa ctggaacgag atcctgccag gctttgaaat caacttcatc 600
aagtctcgga gcagcaacat gctgacgccc tatgactact cctctgtgat gcactatggg 660

aggcttgcct tcagccggcg tgggctgccc accatcacac cactttgggc ccccagtgtc 720
cacatcggcc agcgatggaa cctgagtgcc tcggacatca cccgggtcct ccaactctac 780
ggctgcagcc caaaattaga aaaggtaaat acagttaata taaaaatcat atttctgtac 840
acccagttag aaaacacaat tgttagtaaa cataccatta taatagcaat cataaaggtc 900
ccaagaaata aatctgacag ctgtatcaaa tatttgagga aaaatgaa 948

<210> 17
<211> 2175
<212> DNA
<213> Homo sapiens

<400> 17
ttccagctct ggatatggtt aaggccttgc cctgtgacct ggattcccag gtttcctggt 60
gggggtgtat ttcctggagg cagtctctcc cccttgcaca ttttggggac caaagcattt 120
tttaaagtgc tcttcctgga ccaccatttt agactatata tggagcatag taatgacata 180
atcagtccac attttaaaga gatcactcag attatcacat cttttcaaga aataatagag 240
gaagagtttg gaatatcaca atgctataca tataataatc catctaaatc tgacattagg 300
atccattgga cagtatcaga tttgtctcag gtttttattc taagtagagc caaaagcctt 360
tatatccaaa tcttgtctca aagacattcc aggaaaaaac gtcttatatc atatccaaga 420
tacattgaaa ttatggttac agctgatgct aaagtggttt ctgctcatgg atcgaatttg 480
caaaactata tactgactct aatgtcaatt gttgcaacaa tctacaaaga tccaagtatt 540
ggaaatttga tacacatagt agtggtaaaa ttagttatga ttcaccgtga ggaggaagga 600
ccagtcatta attttgatgg tgctaccaca ttaaagaact tttgttcatg gcaacaaact 660
cagaatgacc ttgatgatgt tcacccttcc caccatgaca ctgctgttct tatcactagg 720
gaagacattt gttcatctaa agagaaatgt aacatgttag gtttatcata tttaggtacc 780
atatgtgatc ctttacaaag ctgctttatt aatgaagaaa aaggactcat ttctgctttt 840
actatagccc atgagcttgg gcacacactt ggtgttcaac atgatgataa tcctagatgt 900
aaagaaatga aagttacaaa gtatcatgta atggcccctg ctttaagttt tcacatgagt 960
ccttggagct ggtcaaactg tagtcggaaa tatgttactg aattcctaga tactggttac 1020
ggggaatgtc ttcttgacaa accagatgaa gaaatatata atctgccttc agaacttcct 1080
ggatcacgat atgatggaaa caagcagtgt gagcttgcgt ttggtcctgg gtcacaaatg 1140
tgtccccata tagagaatat atgcatgcat ctgtggtgca caagcacaga aaagcttcac 1200
aaaggctgtt tcactcaaca cgtgccacca gcagatggaa cagactgcgg tcctggaatg 1260
cattgccgtc atgggctatg tgtaaacaaa gaaacggaaa cacgtcctgt aaatggtgaa 1320
tggggaccat gggaaccttа cagttcttgt tcaagaacat gtggaggcgg aatcgaaagt 1380
gcaaccaggc gctgtaatcg tcctgagcca agaaacggag gaaattactg tgtgggccgc 1440
aggatgaaat ttcgatcatg taatactgat tcatgtccaa aaggcacaca agactttcga 1500
gagaagcagt gctctgattt taatggtaaa catttggaca tcagtggcat tccctctaat 1560
gtgaggtggc ttccaagata cagtggcatt ggcacaaagg atcgttgtaa actctattgt 1620
caggttgctg gaaccaatta tttctaccta ttgaaggata tggttgaaga tggtactcct 1680
tgtggaactg aaactcatga catctgtgtt caaggccagt gtatggcagc tggttgtgat 1740
cacgtgttaa actccagtgc caagatagac aaatgtggag tgtgtggtgg ggacaactct 1800
tcatgcaaga caataacagg tgtcttcaac agttctcatt atggttataa tgttgttgta 1860
aagattcccg caggagcaac aaacgttgac attcgtcagt acagctattc tggacaacca 1920
gatgacagtt accttgcatt atctgacgct gaagggaatt ttcttttcaa tggaaatttt 1980
cttctaagta cgtcaaaaaa agaaatcaat gtgcaaggaa caagaactgt tattgaatac 2040
agtggatcaa ataacgcagt tgaaagaatt aatagtacta atcgacaaga gaaagaactt 2100
attttgcagg tgttgtgtgt gggtaattta tacaaccctg atgtacatta ttccttcaat 2160
atccctttgg aagag 2175

<210> 18
<211> 2100
<212> DNA
<213> Homo sapiens

<400> 18
cccagcctcc tgtcctgcct tcttagcttt ccccgcccag ggccagatat agcctggcag 60
ctgtcttgta aaggaagttg gattggaaca cagacacact cactcacagt ctcatctgcg 120
actgattttg tgcttcggtg gcagaatcga atggttgagt atcccggagt tcctcagatg 180
ccctatggag gccacagtag ccccatgact tttctccttt atggagatat tgccaacttt 240
gatttttaca gcaaccttgt ggtgacagcc cctcctgtgg gttggacctc actctcttcc 300
tgtcttgacc ttcccaacct gctgggcctg gtgggggacc agctgggcga cacagagcgg 360

aagcggcggc atgccaagcc aggcagctac agcatcgagg tgctgctggt ggtggacgac 420
tcggtggttc gcttccatgg caaggagcat gtgcagaact atgtcctcac cctcatgaat 480
atcgtgaatg agatttacca cgatgagtcc ctgggggttc atataaatat tgccctcgtc 540
cgcttgatca tggttggcta ccgacagtcc ctgagcctga tcgagcgcgg gaacccctca 600
cgcagcctgg agcaggtgtg tcgctgggca cactcccagc agcgccagga ccccagccac 660
gctgagcacc atgaccacgt tgtgttcctc acccggcagg actttgggcc ctcaggtatg 720
caagggtatg cacccgtcac tggcatgtgt caccccctga ggagctgtgc cctcaaccat 780
gaggatggct tctcctcagc cttcgtgata gctcatgaga ccggccacgt gctcggcatg 840
gagcatgacg gtcaggggaa tggctgtgca gatgagacca gcctgggcag cgtcatggcg 900
cccctggtgc aggctgcctt ccaccgcttc cattggtccc gctgcagcaa gctggagctc 960
agccgctacc tcccctccta cgactgcctc ctcgatgacc cctttgatcc tgcctggccc 1020
cagcccccag agctgcctgg gatcaactac tcaatggatg agcagtgccg ctttgacttt 1080
ggcagtggct accagacctg cttggcattc aggacctttg agccctgcaa gcagctgtgg 1140
tgcagccatc ctgacaaccc gtacttctgc aagaccaaga aggggccccc gctggatggg 1200
actgagtgtg cacccggcaa gtggtgcttc aaaggtcact gcatctggaa gtcgccggag 1260
cagacatatg gccaggatgg aggctggagc tcctggacca agtttgggtc atgttcgcgg 1320
tcatgtgggg gcggggtgcg atcccgcagc cggagctgca acaacccctc cccagcctat 1380
ggaggccgcc tgtgcttagg gcccatgttc gagtaccagg tctgcaacag cgaggagtgc 1440
cctgggacct acgaggactt ccgggcccag cagtgtgcca agcgcaactc ctactatgtg 1500
caccagaatg ccaagcacag ctgggtgccc tacgagcctg acgatgacgc ccagaagtgt 1560
gagctgatct gccagtcggc ggacacgggg gacgtggtgt tcatgaacca ggtggttcac 1620
gatgggacac gctgcagcta ccgggaccca tacagcgtct gtgcgcgtgg cgagtgtgtg 1680
cctgtcggct gtgacaagga ggtggggtcc atgaaggcgg atgacaagtg tggagtctgc 1740
gggggtgaca actcccactg caggactgtg aaggggacgc tgggcaaggc ctccaagcag 1800
gcagctctca agctggtgca gatcccagca ggtgccaggc acatccagat tgaggcactg 1860
gagaagtccc cccaccgcat tgtggtgaag aaccaggtca ccggcagctt catcctcaac 1920
cccaagggca aggaagccac aagccggacc ttcaccgcca tgggcctgga gtgggaggat 1980
gcggtggagg atgccaagga aagcctcaag accagcgggc ccctgcctga agccattgcc 2040
atcctggtga gccccactct tgatacacaa aatattaaag aacccaggca cagaccagac 2100

<210> 19
<211> 2247
<212> DNA
<213> Homo sapiens

<400> 19
ttgttgttgt tgttgttgtt gttgttgttt ttgagacaga gttggcaggg acctataatc 60
tcagcgactc aggaggctga ggcagtagaa tcacttgaac ccaggaggcg gaggttgcat 120
tcaggagttc gagaccagcc tggtcaacat ggtgaaaccc tgtctctact gaaaatccaa 180
aaattagaca ggcatggttt aataaggaca cgaaaaaatg aattcctcat ctcgccatta 240
cctcagcttc tggcccagga acacaactac agctcccctg cgggtcacca tcctcacgta 300
ctattggatt gcttttatca ttgtcacatc aaagatttct cttcttccct tgtttcagtt 360
tctctctcaa ctgtcttatc cagatattct ctaatatccc ttccaaatgc tcttctgttc 420
atcgtagatg ctcccaagcc tcccacagag gacacctatc taaggtttga tgaatatggg 480
agctctgggc gacccagaag atcagctgga aaatcacaaa agggcctcaa tgtggaaacc 540
ctcgtggtgg cagacaagaa aatggtggaa aagcatggca agggaaatgt caccacatac 600
attctcacag taatgaacat ggtttctggc ctatttaaag atgggactat tggaagtgac 660
ataaacgtgg ttgtggtgag cctaattctt ctggaacaag aacctggagg attattgatc 720
aaccatcatg cagaccagtc tctgaatagt ttttgtcaat ggcagtctgc cctcattgga 780
aagaatggca agagacatga tcatgccatc ttactaacag gatttgatat ttgttcttgg 840
aagaatgaac catgtgacac tctagggttt gcccccatca gtggaatgtg ctctaagtac 900
cgaagttgta ccatcaatga ggacacagga cttggccttg ccttcaccat cgctcatgag 960
tcagggcaca actttggtat gattcacgat ggagaaggga atccctgcag aaaggctgaa 1020
ggcaatatca tgtctcccac actgaccgga aacaatggag tgttttcatg gtcttcctgc 1080
agccgccagt atctcaagaa attcctcagc acacctcagg cggggtgtct agtggatgag 1140
cccaagcaag caggacagta taaatatccg gacaaactac caggacagat ttatgatgct 1200
gacacacagt gtaaatggca atttggagca aaagccaagt tatgcagcct tggttttgat 1260
atttgcaaat cactttggtg ccaccgagta ggccacaggt gtgagaccaa gtttatgccc 1320
gcagcagaag ggaccgtttg tggcttgagt atgtggtgtc ggcaaggcca gtgcgtaaag 1380
tttggggagc tcgggccccg gcccatccac ggccagtggt ccgcctggtc gaagtggtca 1440
gaatgttccc ggacatgtgg tggaggagtc aagttccagg agagacactg caataacccc 1500

aagcctcagt atggtggctt attctgtcca ggttctagcc gtatttatca gctgtgcaat 1560
attaaccctt gcaatgaaaa tagcttggat tttcgggctc aacagtgtgc agaatataac 1620
agcaaacctt tccgtggatg gttctaccag tggaaaccct atacaaaagt ggaagaggaa 1680
gatcgatgca aactgtactg caaggctgag aactttgaat ttttttttgc aatgtccggc 1740
aaagtgaaag atggaactcc ctgctcccca aacaaaaatg atgtttgtat tgacggggtt 1800
tgtgaactag tgggatgtga tcatgaacta ggctctaaag cagtttcaga tgcttgtggc 1860
gtttgcaaag gtgataattc aacttgcaag ttttataaag gcctgttcaa gcaattctcc 1920
tgcctcaccc tcttgaaata ttatccggtg gtcctcattc cagctggcgc ccgaagcatc 1980
gaaatccagg agctgcaggt ttcctccagt tacctcgcag ttcgaagcct cagtcaaaag 2040
tattacctca ccgggggctg gagcatcgac tggcctgggg agttcccctt cgctgggacc 2100
acgtttgaat accagcgctc tttcaaccgc ccggaacgtc tgtacgcgcc agggcccaca 2160
aatgagacgc tgtgttgctc tgtcgcccag gctggagggc agttgcgtga tctcggctca 2220
ctgcaagctc cgcctcctga gttcacg 2247

<210> 20
<211> 2166
<212> DNA
<213> Homo sapiens

<400> 20
ctcctcctct ggcgctgccc gctgtctccc gccttccctc tgctcccctc tcgcctccgc 60
ctcagcgccc cgctgacctc gcctcctccc ctctgctctt tgtccctgca ctctcccctc 120
ctcggtcctc tgaccccccc gccctcacct cctcccctcc tctctcccct gcccgccccg 180
cgctctccca ccgctcccgc cgcccccgcc gccgcggctg ccactccgcc ccccgcgccg 240
cacggagctt cacctctcct cactcttttg atttcagaat atgacctggt ctctgcctac 300
gaggttgacc acaggggcga ttacgtgtcc catgaaatca tgcaccatca gcggcggaga 360
agagcagtgg ccgtgtccga ggttgagtct cttcaccttc ggctgaaagg ccccaggcac 420
gacttccaca tggatctgag gacttccagc agcctagtgg ctcctggctt tattgtgcag 480
acgttgggaa agacaggcac taagtctgtg cagactttac cgccagagga cttctgtttc 540
tatcaaggct ctttgcgatc acacagaaac tcctcagtgg ccctttcaac ctgccaaggc 600
ttgtcaggca tgatacgaac agaagaggca gattacttcc taaggccact tccttcacac 660
ctctcatgga aactcggcag agctgcccaa ggcagctcgc catcccacgt actaaatgaa 720
gaactgaacg tggagacctt ggtggtggtc gacaaaaaga tgatgcaaaa ccatggccat 780
gaaaatatca ccacctacgt gctcacgata ctcaacatgg tatctgcttt attcaaagat 840
ggaacaatag gaggaaacat caacattgca attgtaggtc tgattcttct agaagatgaa 900
cagccaggac tggtgataag tcaccacgca gaccacacct taagtagctt ctgccagtgg 960
cagtctggat tgatggggaa agatgggact cgtcatgacc acgccatctt actgactggt 1020
ctggatatat gttcctggaa gaatgagccc tgtgacactt tgggatttgc acccataagt 1080
ggaatgtgta gtaaatatcg cagctgcacg attaatgaag atacaggtct tggactggcc 1140
ttcaccattg cccatgagtc tggacacaac tttggcatga ttcatgatgg agaagggaac 1200
atgtgcaaaa agtccgaggg caacatcatg tcccctacat tggcaggacg caatggagtc 1260
ttctcctggt caccctgcag ccgccagtat ctacacaaat ttctaagcac cgctcaagct 1320
atctgccttg ctgatcagcc aaagcctgtg aaggaataca agtatcctga gaaattgcca 1380
ggagaattat atgatgcaaa cacacagtgc aagtggcagt tcggagagaa agccaagctc 1440
tgcatgctgg actttaaaaa ggcaaccctg tggtgccatc gtattggaag gaaatgtgag 1500
actaaattta tgccagcagc agaaggcaca atttgtgggc atgacatgtg gtgccgggga 1560
ggacagtgtg tgaaatatgg tgatgaaggc cccaagccca cccatggcca ctggtcggac 1620
tggtcttctt ggtccccatg ctccaggacc tgcggagggg gagtatctca taggagtcgc 1680
ctctgcacca accccaagcc atcgcatgga gggaagttct gtgagggctc cactcgcact 1740
ctgaagctct gcaacagtca gaaatgtccc cgggacagtg ttgacttccg tgctgctcag 1800
tgtgccgagc acaacagcag acgattcaga gggcggcact acaagtggaa gccttacact 1860
caagtagaag atcaggactt atgcaaactc tactgtatcg cagaaggatt tgatttcttc 1920
ttttctttgt caaataaagt caaagatggg actccatgct cggaggatag ccgtaatgtt 1980
tgtatagatg ggatatgtga gagagttgga tgtgacaatg tccttggatc tgatgctgtt 2040
gaagacgtct gtggggtgtg taacgggaat aactcagcct gcacgattca caggggtctc 2100
tacaccaagc accaccacac caaccgtgag tactttagag ctgcctgcaa gccttgggca 2160
aagaag 2166

<210> 21
<211> 2229
<212> DNA
<213> Homo sapiens

<400> 21
atgtcctctt accgtcccag ggtcatatcc agaatgccac atcttctcag gctcctcttg 60
gctgtgacag tttctcagac tttcattgtt ttcttgtttg tgtttttgtt ttttgtgatc 120
ttgacagtct tgaggattac tgatcaggtg agcgaggtct gcaccacccc tggctgcgtg 180
atagcagcag ctgccaggat cctccagaac atggacccga ccacggaacc gtgtgacgac 240
ttctaccagt ttgcatgcgg aggctggctg cggcgccacg tgatccctga gaccaactca 300
agatacagca tctttgacgt cctccgcgac gagctggagg tcatcctcaa agcggtgctg 360
gagaattcga ctgccaagga ccggccggct gtggagaagg ccaggacgct gtaccgctcc 420
tgcatgaacc agagtgtgat agagaagcga ggctctcagc ccctgctgga catcttggag 480
gtggtgggag gctggccggt ggcgatggac aggtggaacg agaccgtagg actcgagtgg 540
gagctggagc ggcagctggc gctgatgaac tcacagttca acaggcgcgt cctcatcgac 600
ctcttcatct ggaacgacga ccagaactcc agccggcaca tcatctacat agaccagccc 660
accttgggca tgccctcccg agagtactac ttcaacggcg gcagcaaccg gaaggtgcgg 720
gaagcctacc tgcagttcat ggtgtcagtg gccacgttgc tgcgggagga tgcaaacctg 780
cccagggaca gctgcctggt gcaggaggac atggtgcagg tgctggagct ggagacacag 840
ctggccaagg ccacggtacc ccaggaggag agacacgacg tcatcgcctt gtaccaccgg 900
atgggactgg aggagctgca aagccagttt ggcctgaagg gatttaactg gactctgttc 960
atacaaactg tgctatcctc tgtcaaaatc aagctgctgc cagatgagga agtggtggtc 1020
tatggcatcc cctacctgca gaaccttgaa aacatcatcg acacctactc agccaggacc 1080
atacagaact acctggtctg gcgcctggtg ctggaccgca ttggtagcct aagccagaga 1140
ttcaaggaca cacgagtgaa ctaccgcaag gcgctgtttg gcacaatggt ggaggaggtg 1200
cgctggcgtg aatgtgtggg ctacgtcaac agcaacatgg agaacgccgt gggctccctc 1260
tacgtcaggg aggcgttccc tggagacagc aagagcatgg tcagagaact cattgacaag 1320
gtgcggacag tgtttgtgga gacgctggac gagctgggct ggatggacga ggagtccaag 1380
aagaaggcgc aggagaaggc catgagcatc cgggagcaga tcgggcaccc tgactacatc 1440
ctggaggaga tgaacaggcg cctggacgag gagtactcca atctgaactt ctcagaggac 1500
ctgtactttg agaacagtct gcagaacctc aaggtgggcg cccagcggag cctcaggaag 1560
cttcgggaaa aggtggaccc aaatctctgg atcatcgggg cggcggtggt caatgcgttc 1620
tactccccaa accgaaacca gattgtattc cctgccggga tcctccagcc ccccttcttc 1680
agcaaggagc agccacaggc cttgaacttt ggaggcattg ggatggtgat cgggcacgag 1740
atcacgcacg gctttgacga caatggccgg aacttcgaca agaatggcaa catgatggat 1800
tggtggagta acttctccac ccagcacttc cgggagcagt cagagtgcat gatctaccag 1860
tacggcaact actcctggga cctggcagac gaacagaacg tgaacggatt caacaccctt 1920
ggggaaaaca ttgctgacaa cggaggggtg cggcaagcct ataaggccta cctcaagtgg 1980
atggcagagg gtggcaagga ccagcagctg cccggcctgg atctcaccca tgagcagctc 2040
ttcttcatca actatgccca ggtgtggtgc gggtcctacc ggcccgagtt cgccatccaa 2100
tccatcaaga cagacgtcca cagtcccctg aaggtactgg ggtcgctgca gaacctggcc 2160
gccttcgcag acacgttcca ctgtgcccgg ggcaccccca tgcaccccaa ggagcgatgc 2220
cgcgtgtgg 2229

<210> 22
<211> 1113
<212> DNA
<213> Homo sapiens

<400> 22
atgactgcac ttgaccgtgc ttgtctttat tggctgtttt tatttaaatt attagttatt 60
gatataaaaa ataatggtca cttttatgta actctcgcca actcaaaaca tcttagcctg 120
gactttattg tccatatcac tatcagtatt ttggtcaaag ccattcaacg agtctctagg 180
aagttccaaa cttttccaca ttttcctgtc ttctactggg cccttcaaac tatctatgag 240
tggatgagag agatcagtga gaagtacaag gaagtggtga cacagcattt cctaggagtg 300
acctatgaga cccaccccat gtattatctg aagatcagcc aaccatctgg taatcccaag 360
aaaatcattt ggatggactg tggaattcac gccagagaat ggattgctcc tgctttttgc 420
caatggttcg tcaaagaaat tctacaaaac cataaagaca actcaagtat acgcaagctc 480
cttaggaacc tggacttcta tgtccttcca gttcttaaca tagatggtta tatctacact 540
tggacaactg atcgtctttg gaggaaatcc cgttcacccc ataataatgg cacatgtttt 600
gggacggatc tcaatcgaaa tttcaatgca tcttggtgta gtattggtgc ctctagaaac 660
tgccaagatc aaacattctg tgggacaggg ccagtgtctg aaccagagac taaagctgtt 720
gccagcttca tagagagcaa gaaggatgat attttgtgct tcctgaccat gcactcttat 780
gggcagttaa ttctcacacc ttacggctac accaaaaata aatcaagtaa ccacccagaa 840
atgattcaag ttggacagaa ggcagcaaat gcattgaaag caaagtatgg aaccaattat 900
agagttggat cgagtgcaga tattttatat gcctcatcag ggtcttcaag agattgggcc 960

cgagacattg ggattccctt ctcatatacg tttgagctga gggacagtgg aacatatggg 1020
tttgttctgc cagaagctca gatccagccc acctgtgagg agaccatgga ggctgtgctg 1080
tcagtcctgg atgatctaca aaaaaaccca tat 1113

<210> 23
<211> 954
<212> DNA
<213> Homo sapiens

<400> 23
cctaacacac aaaaccatat gcctctgtgt ctagagttag gcatcaggag ttatcactct 60
gggttttgcc aagattgctt tagaaggaat gaagatattt cacacagtat agttttgccg 120
gctgcagttt cttcagctca tccggttcct aagcacataa agaagccaga ctatgtgacg 180
acaggcattg taccagactg gggagacagc atagaagtta agaatgaaga tcagattcaa 240
gggcttcatc aggcttgtca gctggcccgc cacgtcctcc tcttggctgg gaagagttta 300
aaggttgaca tgacaactga agagatagat gctcttgttc atcgggaaat catcagtcat 360
aatgcctatc cctcacctct aggctatgga ggttttccaa aatctgtttg tacctctgta 420
aacaacgtgc tctgtcatgg tattcctgac agtcgacctc ttcaggatgg agatattatc 480
aacattgatg tcacagtcta ttacaatggc taccatggag acacctctga aacatttttg 540
gtgggcaatg tggacgaatg tggtaaaaag ttagtggagg ttgccaggag gtgtagagat 600
gaagcaattg cagcttgcag agcaggggct cccttctctg taattggaaa cacaatcagc 660
cacataactc atcagaatgg ttttcaagtc tgtccacatt ttgtgggaca tggaatagga 720
tcttactttc atggacatcc agaaatttgg catcatgcaa acgacagtga tctacccatg 780
gaggagggca tggcattcac tatagagcca atcatcacgg agggatcccc tgaatttaaa 840
gtcctggagg atgcatggac tgtggtctcc ctagacaatc aaaggtcggc gcagttcgag 900
cacacggttc tgatcacgtc gaggggcgcg cagatcctga ccaaactacc ccat 954

<210> 24
<211> 711
<212> DNA
<213> Homo sapiens

<400> 24
aggatcgtgg gcggcatgga agcatccccg ggggagtttc cgtggcaagc cagccttcga 60
gagaacaagg agcacttctg tggggccgcc atcatcaacg ccaggtggct ggtgtctgct 120
gctcactgct tcaatgagtt ccaagacccg acgaagtggg tggcctacgt gggtgcgacc 180
tacctcagcg gctcggaggc cagcaccgtg cgggcccagg tggtccagat cgtcaagcac 240
cccctgtaca acgcggacac ggccgacttt gacgtggctg tgctggagct gaccagccct 300
ctgcctttcg gccggcacat ccagcccgtg tgcctcccgg ctgccacaca catcttccca 360
cccagcaaga agtgcctgat ctcaggctgg ggctacctca aggaggactt cctggtcaag 420
ccagaggtgc tgcagaaagc cactgtggag ctgctggacc aggcactgtg tgccagcttg 480
tacggccatt cactcactga caggatggtg tgcgctggct acctggacgg gaaggtggac 540
tcctgccagg gtgactcagg aggacccctg gtctgcgagg agccctctgg ccggttcttt 600
ctggctggca tcgtgagctg gggaatcggg tgtgcggaag cccggcgtcc aggggtctat 660
gcccgagtca ccaggctacg tgactggatc ctggaggcca ccaccaaagc c 711

<210> 25
<211> 714
<212> DNA
<213> Homo sapiens

<400> 25
cgcattgttg gtggagctgt gtcctccgag ggtgagtggc catggcaggc cagcctccag 60
gttcggggtc gacacatctg tgggggggcc ctcatcgctg accgctgggt gataacagct 120
gcccactgct tccaggagga cagcatggcc tccacggtgc tgtggaccgt gttcctgggc 180
aaggtgtggc agaactcgcg ctggcctgga gaggtgtcct tcaaggtgag ccgcctgctc 240
ctgcacccgt accacgaaga ggacagccat gactacgacg tggcgctgct gcagctcgac 300
cacccggtgg tgcgctcggc cgccgtgcgc cccgtctgcc tgcccgcgcg ctcccacttc 360
ttcgagcccg gcctgcactg ctggattacg ggctggggcg ccttgcgcga gggcggcccc 420
atcagcaacg ctctgcagaa agtggatgtg cagttgatcc cacaggacct gtgcagcgag 480
gtctatcgct accaggtgac gccacgcatg ctgtgtgccg gctaccgcaa gggcaagaag 540
gatgcctgtc agggtgactc aggtggtccg ctggtgtgca aggcactcag tggccgctgg 600

ttcctggcgg ggctggtcag ctggggcctg ggctgtggcc ggcctaacta cttcggcgtc 660
tacacccgca tcacaggtgt gatcagctgg atccagcaaa ccatggccca gagc 714

<210> 26
<211> 705
<212> DNA
<213> Homo sapiens

<400> 26
aggattgtgg gcggcagcgc agcgggccgt ggggagtggc cgtggcaggt gagcctgtgg 60
ctgcggcgcc gggaacaccg ttgcggggcc gtgctggtgg cagagaggtg gctgctgtcg 120
gcggcgcact gcttcgacgt ctacggggac cccaagcagt gggcggcctt cctaggcacg 180
ccgttcctga gcggcgcgga ggggcagctg gagcgcgtgg cgcgcatcta caagcacccg 240
ttctacaatc tctacacgct cgactacgac gtggcgctgc tggagctggc ggggccggtg 300
cgtcgcagcc gcctggtgcg tcccatctgc ctgcccgagc ccgcgccgcg acccccggac 360
ggcacgcgct gcgtcatcac cggctggggc tcggtgcgcg aaggaggctc catggcgcgg 420
cagctgcaga aggcggccgt gcgcctcctc agcgagcaga cctgccgccg cttctaccca 480
gtgcagatca gcagccgcat gctgtgtgcc ggcttcccgc agggtggcgt ggacagctgc 540
tcgggtgacg ctgggggacc cctggcctgc agggagccct ctggacggtg ggtgctaact 600
ggggtcacta gctggggcta tggctgtggc cggccccact tcccaggtgt ctatacccgg 660
gtggcagctg tgagaggctg gataggacag cacatccagg acaac 705

<210> 27
<211> 714
<212> DNA
<213> Homo sapiens

<400> 27
cgcatcatcg gaggcacaga caccctggag gggggttggc cgtggcaggt cagcctccac 60
tttgttggat ctgcctactg tggtgcctca gtcatctcca gggagtggct tctttctgca 120
gcccactgtt ttcatggaaa caggctgtca gatcccacac catggactgc acacctcggg 180
atgtatgttc aggggaatgc caagtttgtc tccccggtga gaagaattgt ggtccacgag 240
tactataaca gtcagacttt tgattatgat attgctttgc tacagctcag tattgcctgg 300
cctgagaccc tgaaacagct cattcagcca atatgcattc ctcccactgg tcagagagtt 360
cgcagtgggg agaagtgctg ggtaactggc tgggggcgaa gacacgaagc agataataaa 420
ggctccctcg ttctgcagca agcggaggta gagctcattg atcaaacgct ctgtgtttcc 480
acctacggga tcatcacttc tcggatgctc tgtgcaggca taatgtcagg caagagagat 540
gcctgcaaag gagattcggg tggaccttta tcttgtcgaa gaaaaagtga tggaaaatgg 600
attttgactg gcattgttag ctggggacat ggaagtggac gaccaaactt tcctggtgtt 660
tacacaaggg tgtcaaactt tgttccctgg attcataaat atgtcccttc tctt 714

<210> 28
<211> 705
<212> DNA
<213> Homo sapiens

<400> 28
agaattgtcc aaggaaggga aacagctatg gaaggggaat ggccatggca ggccagcctc 60
cagctcatag ggtcaggcca tcagtgtgga gccagcctca tcagtaacac atggctgctc 120
acagcagctc actgcttttg gaaaaataaa gacccaactc aatggattgc tacttttggt 180
gcaactataa caccacccgc agtgaaacga aatgtgagga aaattattct tcatgagaat 240
taccatagag aaacaaatga aaatgacatt gctttggttc agctctctac tggagttgag 300
ttttcaaata tagtccagag agtttgcctc ccagactcat ctataaagtt gccacctaaa 360
acaagtgtgt tcgtcacagg atttggatcc attgtagatg atggacctat acaaaataca 420
cttcggcaag ccagagtgga aaccataagc actgatgtgt gtaacagaaa ggatgtgtat 480
gatggcctga taactccagg aatgttatgt gctggattca tggaaggaaa aatagatgca 540
tgtaagggag attctggtgg acctctggtt tatgataatc atgacatctg gtacattgta 600
ggtatagtaa gttggggaca atcatgtgca cttcccaaaa aacctggagt ctacaccaga 660
gtaactaagt atcgagattg gattgcctca aagactggta tgaac 705

<210> 29
<211> 690

<212> DNA
<213> Homo sapiens

<400> 29
cgcatcgtgg ggggcagcgc ggcgccgccc ggggcctggc cctggctggt gaggctgcag 60
ctcggcgggc agcctctgtg cggcggcgtc ctggtagcgg cctcctgggt gctcacggca 120
gcgcactgct ttcttctgtg gactgtgacg ctggcagagg ggtcccgggg ggagcaagcg 180
gaggaggtgc cagtgaaccg catcctgccc caccccaagt ttgacccgcg gaccttccac 240
aacgacctgg ccctggtgca gctgtggacg ccggtgagcc cgggggatc ggcgcgcccc 300
gtgtgcctgc cccaggagcc ccaggagccc cctgccggaa ccgcctgcgc catcgcgggc 360
tggggcgccc tcttcgaaga cgggcctgag gctgaagcag tgagagaggc ccgtgttccc 420
ctgctcagca ccgacacctg ccgaagagcc ctggggcccg ggctgcgccc cagcaccatg 480
ctctgcgccg ggtacctggc gggggggcgtt gactcgtgcc agggtgactc gggaggcccc 540
ctgacctgtt ctgagcctgg cccccgccct agagaggtcc tgttcggagt cacctcctgg 600
ggggacggct gcggggagcc agggaagccc ggggtctaca cccgcgtggc agtgttcaag 660
gactggctcc aggagcagat gagcggtgag 690

<210> 30
<211> 699
<212> DNA
<213> Homo sapiens

<400> 30
agaatagcat ctggagtcat tgcacccaag gcggcctggc cttggcaagc ttcccttcag 60
tatgataaca tccatcagtg tggggccacc ttgattagta acacatggct tgtcactgca 120
gcacactgct tccagaagta taaaaatcca catcaatgga ctgttagttt tggaacaaaa 180
atcaaccctc ccttaatgaa aagaaatgtc agaagattta ttatccatga gaagtaccgc 240
tctgcagcaa gagagtacga cattgctgtt gtgcaggtct cttccagagt caccttttcg 300
gatgacatac gccagatttg tttgccagaa gcctctgcat ccttccaacc aaatttgact 360
gtccacatca caggatttgg agcactttac tatggtgggg aatcccaaaa tgatctccga 420
gaagccagag tgaaaatcat aagtgatgat gtctgcaagc aaccacaggt gtatggcaat 480
gatataaaac ctggaatgtt ctgtgccgga tatatggaag gaatttatga tgcctgcagg 540
ggtgattctg ggggaccttt agtcacaagg gatctgaaag atacgtggta tctcattgga 600
attgtaagct ggggagataa ctgtggtcaa aaggacaagc ctggagtcta cacacaagtg 660
acttattacc gaaactggat tgcttcaaaa acaggcatc 699

<210> 31
<211> 741
<212> DNA
<213> Homo sapiens

<400> 31
cggattatag ggggcaccga agcacaagct ggcgcatggc cgtgggtggt gagcctgcag 60
attaaatatg gccgtgttct tgttcatgta tgtgggggaa ccctagtgag agagaggtgg 120
gtcctcacag ctgcccactg cactaaagac gctagcgatc ctttaatgtg gacagctgtg 180
attggaacta ataatataca tggacgctat cctcatacca agaagataaa aattaaagca 240
atcattattc atccaaactt cattttggaa tcttatgtaa atgatattgc acttttcac 300
ttaaaaaaag cagtgaggta taatgactat attcagccta tttgcctacc ttttgatgtt 360
ttccaaatcc tggacggaaa cacaaagtgt tttataagtg gctggggaag aacaaaagaa 420
gaaggtaatt atggtaacgc tacaaatatt ttacaagatg cagaagtgca ttatatttct 480
cgagagatgt gtaattctga gaggagttat gggggaataa ttcctaacac ttcattttgt 540
gcaggtgatg aagatggagc ttttgatact tgcaggggtg acagtggggg accattaatg 600
tgctacttac cagaatataa aagatttttt gtaatgggaa ttaccagtta cggacatggc 660
tgtggtcgaa gaggttttcc tggtgtctat attgggccat ccttctacca aaagtggctg 720
acagagcatt tcttccatgc a 741

<210> 32
<211> 759
<212> DNA
<213> Homo sapiens

<400> 32

agaattagta gttggagaaa ttcaacagtg actggacatc catggcaggt ctccctaaaa 60
tcagatgagc accacttctg tggaggaagc ttgattcaag aagatcgggt tgttacagca 120
gcacactgcc tggacagcct cagtgagaag cagctgaaga atataactgt gacttctggg 180
gagtacagcc tctttcagaa ggataagcaa gaacagaata ttcctgtctc aaaaattatt 240
acccatcctg aatacaacag ccgtgaatat atgagtcctg atattgcact gctgtatcta 300
aaacacaaag tcaagtttgg aaatgctgtt cagccaatct gtcttcctga cagcgatgat 360
aaagttgaac caggaattct ttgcttatcc agtggatggg gcaagatttc caaaacatca 420
gaatattcaa atgtcctaca agaaatggaa cttcccatca tggatgacag agcgtgtaat 480
actgtgctca agagcatgaa cctccctccc ctgggaagga ccatgctgtg tgctggcttc 540
cctgattggg gaatggacgc ctgccagggg gactctggag gaccactggt ttgtagaaga 600
ggtggtggaa tctggattct tgctgggata acttcctggg tagctggttg tgctggaggt 660
tcagttcccg taagaaacaa ccatgtgaag gcatcacttg gcattttctc caaagtgtct 720
gagttgatgg attttatcac tcaaaacctg ttcacaggt 759

<210> 33
<211> 723
<212> DNA
<213> Homo sapiens

<400> 33
gagatttggt caggggaaca ggggcagaat gatatggttt ggctctcaag ccttaaaatg 60
agtgggcaac actactgtgg ggcatcattg atcagtgaaa gacacttggt gactgcagct 120
cactgtttta aagtgacaaa aaatccaaaa aactatactg tcagctttgg cacgaaagta 180
actcttccct atatgcaaca tgatgttcaa caaattatta ttcatgaaga ctacatccag 240
gatgaacatc atgatgatat cgcacttata ctgctcacta aaaaagtttt atttaagaat 300
gatgtacatc gagtttgtct tcctgaagcc acacagattt ttccacctgg tgaaggagtt 360
gttgttacag gatggggaag actttcattt aatggtaaga tcagtgaaaa cttaacatac 420
cataaagcat ctgtgaagat tactgataca aacacttgta atgctaaaga agcctatcgt 480
agtatggtac aggatagagt gctatgtgct gggtacatgg aaggaaatat agacgcctgc 540
cagggagact ctggaggacc actagttcat cctaattctc taaatatttg gtatatttgg 600
taccttgttg gagtagtgag ctggggaagg aatgaatgtg gtgcaatcaa tagtccaggg 660
gtctacacac agacagatgt ctttttttct ttaaagtgga tcaaaagcac aattgctctc 720
aaa 723

<210> 34
<211> 693
<212> DNA
<213> Homo sapiens

<400> 34
aggatagtca gcatggaatc taagaaggga aaagtccaat ggctagtggt cctgtttggc 60
agctcttcca ttcagggaag caggaaagat aaggccataa agacctggac tacttttca 120
tatactgtgt ggctaggatc gattacagta ggtgactcaa ggaaacgtgt gaagtactac 180
gtgtccaaaa tcgtcatcca tcccaagtac caagatacaa cggcagacgt cgccttgttg 240
aaactgtcct ctcaagtcac cttcacttct gccatcctgc ctatttgctt gcccagtgtc 300
acaaagcagt tggcaattcc acccttttgt tgggtgaccg gatggggaaa agttaaggaa 360
agttcagata gagattacca ttctgccctt caggaagcag aagtacccat tattgaccgc 420
caggcttgtg aacagctcta caatcccatc ggtatcttct tgccagcact ggagccagtc 480
atcaaggaag acaagatttg tgctggtgat actcaaaaca tgaaggatag ttgcaagggt 540
gattctggag ggcctctgtc gtgtcacatt gatggtgtat ggatccagac aggagtagta 600
agctggggat tagaatgtgg taaatctctt cctggagtct acaccaatgt aatctactac 660
caaaaatgga ttaatgccac tatttcaaga gcc 693

<210> 35
<211> 669
<212> DNA
<213> Homo sapiens

<400> 35
ttggccttta atccagatta cacagtcagc tccactcccc cttacttggt ctatttgaaa 60
tctgactact tgccctgcgc tggagtcctg atccacccgc tttgggtgat cacagctgca 120
cactgcaatt taccaaagct tcgggtgata ttgggggtta caatcccagc agactctaat 180

gaaaagcatc tgcaagtgat tggctatgag aagatgattc atcatccaca cttctcagtc 240
acttctattg atcatgacat catgctaatc aagctgaaaa cagaggctga actcaatgac 300
tatgtgaaat tagccaacct gccctaccaa actatctctg aaaataccat gtgctctgtc 360
tctacctgga gctacaatgt gtacaaagag cccgattcac tgcaaactgt gaacatctct 420
gtaatctcca agcctcagtg tcgcgatgcc tataaaacct acaacatcac ggaaaatatg 480
ctgtgtgtgg gcattgtgcc aggaaggagg cagccctgca aggaagtttc tgctgccccg 540
gcaatctgca atgggatgct tcaaggaatc ctgtcttttg cggatggatg tgttttgaga 600
gccgatgttg gcatctatgc caaaattttt tactatatac cctggattga aaatgtaatc 660
caaaataac 669

<210> 36
<211> 669
<212> DNA
<213> Homo sapiens

<400> 36
aggtgggcag ccggggtgag ggtgccggcc cagcattcag aggagcctcc ccacaacagg 60
tccactaacc catctgatta ccggatcctg cttgggtatg accagcaaag ccatcccaca 120
gagcacagca agcagatgac agtgaataag atcatggtgc acgctgacta taacgagttg 180
caccgcatgg ggagtgacat caccctgctg cagctgcacc atcatgtgga attcagctcc 240
cacatcctcc ccgcctgcct tccggaacca accacgtggc tggcccctga cagctcctgc 300
tggatatctg gttggggaat ggtcaccgag gatgtcttcc tgcctgagcc cttccaactt 360
caggaggcag aggtcggtgt catggacaac actgtctgcg gatccttttt ccagccccag 420
taccccggcc agccaagcag cagtgactac accatccacg aggacatgct gtgcgctggg 480
gacctcataa caggaaaggc catttgccga cgagactcca ggggtccccc cgtctgccca 540
ttaaatggca cctggttcct gatggggctg tctagttgga gcctcgactg ctgctcaccc 600
gtcggtccca gggtcttcac caggctcccc tacttcacca actggatcag ccagaagaag 660
agggagagc 669

<210> 37
<211> 609
<212> DNA
<213> Homo sapiens

<400> 37
agagttgttt ctggatactt ttcagcaaac atggtttcta ctccctggag aacaggcatt 60
ttacatttta accactgcat tcatgatctg agccaaacag tcctggggga tcatttagtt 120
aaattccatc atactataaa gattatttgc catatattag atcatgctgt ggccctttg 180
tttttgcaaa tttcttccat ttggaatggg aacatttacc caatacctct acctgcattt 240
gtttcctaca agaatgctag tatttgtagg atcatgttgt ggggacatgc tggggacatg 300
cttttcccca tgaactttcc cttgtgtgcc cgcgtggaca gacaacaggg ggagcagtgc 360
gagcacaccg agtttggcta ccaacccgaa accatcaaga atgacatgct gtgcgccggc 420
ttcgaggagg gcaagaagga tgcctgcaag ggcgactcgg gcggcccccc ggtgtgcctc 480
gtgggtcagt cgtggctgca ggcgggggtg atcagctggg gtgagggctg tgcccgccag 540
aaccgcccag gtgtctacat ccgtgtcacc gcccaccaca actggatcca tcggatcatc 600
cccaaactg 609

<210> 38
<211> 705
<212> DNA
<213> Homo sapiens

<400> 38
catataatca atggtaaaag acagatagct ttcccccgaa gaccaggaac aagagaagga 60
tgtccacttt tgctatttct atccaatgca cactgcactc cgccatgggc aacagagcaa 120
gactccaact caaaaaaaaa aaaaaaaaaa gagacagaga aaacaattcc aaaagctaca 180
gttatcaaaa cagatggcca ctataaagaa aacaaaaaca gaaaacatca agtgttggca 240
aagatgtgga gaaattggaa cctttatgca ctgttggttt tctgcaagat taaacataga 300
attactgagc caggcagggt ggctcacgcc tgtaatccca gcactttggg aggccgaggc 360
gggtggatca cgagatgggg gtctcactat gttgcccagg ctggtgagac ctcagacgag 420
ctgcaggaga tgcagctccc gctgatcctg gagccctggt gccacctgct ctacggacac 480
atgtcctaca tcatgcccga catgctgtgt gctggggaca tcctgaatgc taagaccgtg 540

tgtgagggcg actccggggg cccacttgtc tgtgaattca accgcagctg gttgcagatt 600
ggaattgtga gctggggccg aggctgctcc aaccctctgt accctggagt gtatgccagt 660
gtttcctatt tctcaaaatg gatatgtgat aacatagaaa tcacg 705

<210> 39
<211> 684
<212> DNA
<213> Homo sapiens

<400> 39
agggtctctg gaggtaggga cagtgtccca tctttggtac catccaccaa tgcctacaac 60
aggaagaggc ctgagaaccc tcacatgtgt ggaggtttcc tggcctcaaa cattgagcac 120
ctgctgtgtg ctaggcacag gattcaaaaa tccatgacgt ctgctcatag gtcaaaggtt 180
aggagacttg aatctcattg gtacaaaggg aaaagaaaga caaggagtaa agagaaaagg 240
aaaatatttg gaaaatacac cagcaacata aattacgaca taagtctgct gggtttggcc 300
agtcctgccg tcatcactga caaagtaatc ccagcttgtc tgccatcccc aaattatgtg 360
gtcgccgacc agactgaatg ttacatcact gactggggag aaacccaagg tacctttggg 420
gctggctttc tcaaggaagc ccagctccct gtgattgaga atgaagtgtg caatcgctat 480
gagtttctga atggaagagt caaatccact gagctctgtg ctgggcattt ggctggaggc 540
attgacagtt gcaaggtaag aaaagatcaa gagaccaaag ttagtctttt tggtatagga 600
tgtggagatt gggttaggtc cccacatttt tatacataca tacacacata cacaccgtcc 660
attcaagaaa atatcaaaga aaat 684

<210> 40
<211> 756
<212> DNA
<213> Homo sapiens

<400> 40
agaaagctag gcatcttaaa ccaccaggta ctattttggt ataatctatc acttctgtta 60
cattttattg gatataaatc atattccgaa ccgctggcgc tgtttggtga ggatgatgac 120
atggatcccc gtccatcacg cagctatcag gtggcaaatg gtatcgcggt cttgccggtt 180
tccggcacgc tggtcagtaa aacccgtgcg cttcagcctt attccgggat gacgggttac 240
aacgggatca ttgctcgcct gcagcaggca atcagtgacc ccggcgttga cggcattctt 300
ctggatatgg atacgccggg tggaatggtg tccggggcgt ttgactgcgc cgacattatt 360
gcccgtatgc gcgatatcaa acccatctgg gcgctggcca atgacatgaa ctgcagtgca 420
ggtcagctta ttgccagttc ggcatcgcga cggctggtca cacaaacggc cagaaccggc 480
tccattgggg tcatgatggc gcacagtaac tatggcgctg cgctcaaaac taacggcggg 540
cacatgcaca catatgttta ttgcagcact attcacaata gcaaagactt aaaaccaacc 600
caaatgccca tcaataatag actggataaa gaaaatgtgg cacatataca ccatggaata 660
ctgtgcagtc ataaaaagga tgagttcatg tcctttgcag gaacgtggat gaagctagaa 720
accatcattc tcagcaaact aacacaggaa cagaaa 756

<210> 41
<211> 897
<212> DNA
<213> Homo sapiens

<400> 41
atgctggggg tgcttcttca aatctggagg gggagctgga agaagcagac acaggcccag 60
ggccggaggg agagaagccg ccaggcagcc ggggctgtga gtgccggagg ccgacgcgca 120
ctgttattgt atcttagagc tgaactggaa gacaaactgg cctgtgtgga cagcaggctt 180
agactggtga tgaggggggct tgtcctgggg agggcctcag gcagctctgt tagacccaaa 240
cttccgaaag atgtgcgtgc tgatttccag acgcgtatcg atgccactcg tcagatgttt 300
gccgaaaagg tttccgctta taccggcatg tctgttcagg acgtgctgga caccgaagcg 360
gcagtattct ccggccagga atctttggat aacgggctgg cggatgaact tgttaacaat 420
accgatgcgc tcggcgtgat gcgcgaagca ctcgacagac gcaaaaaaac aacccttgga 480
ggaactatgc catcaccttc tgcatcagct gtgaccacta agccagttga ccaggcagca 540
actcagacaa ctgcatcagc tgaacaggcc actaccgttg acacgacaat tgcttccgta 600
gcagcccctg tagatgtcag tgcgcaggtt actgcagcag tagctgcaga gaatagtcgc 660
atcatgggca tcctgaactg cgacgaggct aaagggcgtg agtcacaggc gcgagcactg 720
gccgaaacgc cgggtatgac ggtagagagc gcacagcgca ttctggctgc tgcaccgcaa 780

agtgcccaga tgcgtaccga tacggcgctg gatcgtttga tggaaacagc acccggtgca 840
ctccaagcag gtagcgcatc ttctgatgcc gctgacgatt tgttaaacac ccccgtt 897

<210> 42
<211> 1389
<212> DNA
<213> Homo sapiens

<400> 42
acggacccct ggttctccaa gcagtggtac atgaacagcg aggcccaacc agacctgagc 60
atcctgcagg cctggagtca ggggctgtca ggccagggca tcgtggtctc tgtgctggac 120
gatggcatcg agaaggacca cccggacctc tgggccaact acgacccccc ggccagctat 180
gacttcaatg actacgaccc ggaccccccag ccccgctaca cccccagcaa agagaaccgg 240
cacgggaccc gctgtgctgg ggaggtggcc gcgatggcca acaatggctt ctgtggtgtg 300
ggggtcgctt tcaacgcccg aatcggaggc gtacggatgc tggacggtac catcaccgat 360
gtcatcgagg cccagtcgct gagcctgcag ccgcagcaca tccacattta cagcgccagc 420
tggggtcccg aggacgacgg ccgcacggtg gacggccccg gcatcctcac ccgcgaggcc 480
ttccggcgtg gtgtgaccaa gggccgcggc gggctgggca cgctcttcat ctgggcctcg 540
ggcaacggcg gcctgcacta cgacaactgc aactgcgacg gctacaccaa cagcatccac 600
acgctttccg tgggcagcac cacccagcag ggccgcgtgc cctggtacag cgaagcctgc 660
gcctccaccc tcaccaccac ctacagcagc ggcgtggcca ccgaccccca gatcgtcacc 720
acggacctgc atcacgggtg cacagaccag cacacgggca cctcggcctc agccccactg 780
gcggccggca tgatcgccct agcgctggag gccaacccgt tcctgacgtg gagagacatg 840
cagcacctgg tggtccgcgc gtccaagccg gcgcacctgc aggccgagga ctggaggacc 900
aacggcgtgg ggcgccaagt gagccatcac tacggatacg ggctgctgga cgccgggctg 960
ctggtggaca ccgcccgcac ctggctgccc acccagccgc agaggaagtg cgccgtccgg 1020
gtccagagcc gccccacccc catcctgccg ctgatctaca tcagggaaaa cgtatcggcc 1080
tgcgccggcc tccacaactc catccgctcg ctggagcacg tgcaggcgca gctgacgctg 1140
tcctacagcc ggcgcggaga cctggagatc tcgctcacca gccccatggg cacgcgctcc 1200
acactcgtgg ccatacgacc cttggacgtc agcactgaag gctacaacaa ctgggtcttc 1260
atgtccaccc acttctggga tgagaaccca cagggcgtgt ggaccctggg cctagagaac 1320
aagggctact atttcaacac gggtgagggc ggggcggggc tgtggtgggc ggggcttggc 1380
tctccaacc 1389

<210> 43
<211> 675
<212> DNA
<213> Homo sapiens

<400> 43
atggcgtctc gatatgacag ggcgatcact gtcttctccc cagacggaca ccttttttcaa 60
gttgaatatg cccaggaagc ggtgaagaaa ggatccaccg cggtcggaat tcgaggtacc 120
aatatagttg ttcttggggt agaaaaaaaa tctgttgcca agcttcaaga tgaaagaact 180
gtgaggaaaa tttgtgccct tgatgaccat gtctgcatgg cttttgcagg acttactgct 240
gatgctagag tagtaataaa cagagcccgt gtggagtgcc agagccataa gcttacggtt 300
gaggacccag tcactgtaga atacataact cgcttcatag caactttaaa gcagattaat 360
acaaagagtt atttgaagtt ttccagagaa gtacctttttt tgttttgttt tttgtttttt 420
agctgggatt accggcacat gccaccacac ctggctaact tttttgcagg atacaaaatc 480
aacaaacaaa aatttgcagc atttctatat gccaacaatg aacaatctga aaaagaaatc 540
aagaaagtaa tcccatttat gatagctaca aataaaatta aatgcataga aataaactta 600
accaaagaag tgaaagattt ccacaatgaa aactataaaa cactgatgca agaaactgaa 660
gcagacacca aaaaa 675

<210> 44
<211> 684
<212> DNA
<213> Homo sapiens

<400> 44
tcaaaaggag gaatttcagt gggtctctgt gttcgggatg gggtggtggt ggtaagtaga 60
gatactaaca gccctcacag agttactcct ctgctaaatg aactaatgtg tctcaggtgt 120
tctgggctgg cagcagctgc gaagatggtt gcagcattca tctctctgag gagatcagca 180

gagataaata agtatgttat atatccaaga gatgtatgca ccccttatat agtgaacaga 240
atgtccttga taaaaataaa atatacccaa agcaatggac gaagaccttt tggtatttct 300
gccttaattg taggttttga tgatgatggt atctcaagat tgtatcagac agatccttct 360
ggtacttatc atgcttggaa ggcaaatgca ataggccgaa gtgctaaaac tgttcgagaa 420
tttctagaaa agaattacac agaagatgcc atagcaagtg acagtgaagc tatcaagtta 480
gcaataaaag ctttgctaga agttgtccag tctggtggaa aaaacattga acttgctata 540
ataagaagaa atcaaccttt gaagaagaaa gaggaggagg aggaaagaag gaagaagaaa 600
gaagaagaag gaggagaaga agaagaggag gaggaggaag aggatgagga ggaagaggag 660
gaggtggaag aggaagagga agaa 684

<210> 45
<211> 3015
<212> DNA
<213> Homo sapiens

<400> 45
gagaatggaa gtcttacatg gcaagaatta ctcagacaaa cagggaaatg ctcaataccc 60
tgtctaatcg atacgggcgc tcaagcaaat attataacag aagaaactgt tcgagcacat 120
aaactgccta ccagaccctg gtcaaaaagt gtgatatatg gtggagttta tccaaataag 180
attaatcgca aaacaataaa acttaacata agtctaaatg gaatatcaat caaaacagaa 240
ttcttggttg taaagaaatt ttcgcatcca gctgctatct ccttcacaac attatatgac 300
aataacattg aaatatctag cagtaaacac acgctctctc aaatgaacaa agtttcaaat 360
attgtcaagg aacctgagtt accagatatc tataaagaat tcaaagacat tactgcagaa 420
accaatacgg aaaagctacc aaagccaata aaagggttag aatttgaagt tgaactaact 480
caagaaaact acagattacc tatcagaaat tacccgctac caccgggaaa aatgcaagct 540
atgaatgatg aaattaatca aggattaaaa agtggaatta tacgagaatc taaagccatt 600
aacgcctgtc cagtaatgtt cgttccgaaa aaggaaggca ccttgagaat ggtggttgac 660
tacaaacctt taaataagta tgtcaaaccc aatatatatc cgttaccact tattgaacaa 720
ttacttgcta aaatacaagg ttctacaatt tttactaaac ttgacctcaa aagtgcctat 780
cacttgatac gagtaagaaa aggagatgaa cataaacttg cttttcgctg tcctcgtgga 840
gtttttgaat atctagtaat gccttatggc atatctatag ctccagcaca ttttcaatac 900
tttatcaata caatacttgg tgaagtcaaa gaatcgcatg tagtatgtta tatggataat 960
attttaattc attcaaaatc ggaatctgaa catgtaaaac atgttaaaga cgttctacag 1020
aaattaaaaa atgcgaactt aattatcaat caagcaaaat gtgaatttca tcaatcacaa 1080
gtaaaattta tagggtatca catttcggaa aaaggattta cgccttgtca agaaaatata 1140
gacaaagtct tacaatggaa gcaacctaag aatcgtaaag aattacgaca atttctaggt 1200
tctgtcaatt atcttaggaa attcattcca aagacatcac aattaacaca tccactcaat 1260
aatcttttga aaaaggatgt acgctggaaa tggacaccaa cacaaaccca agcgatagaa 1320
aacattaaac aatgtttagt ttcgcctccg gtgctacgac actttgattt cagtaaaaag 1380
attctactag aaactgatgc ttcagatgtc gctgtaggag ccgtattgtc tcaaaaacat 1440
gatgatgata aatactatcc tgttggatac tattcagcaa agatgtctaa agcacaatta 1500
aattatagcg tatcggacaa agaaatgctt gcaatcatta agtctctcaa acattggaga 1560
cactatttag aatccactat cgaacctttc aaaattttaa cagaccatcg aaacttaatt 1620
ggtcgcatta ctaacgaatc cgagcctgaa aacaaacgtt tagctcgttg gcaattattt 1680
ttacaagact tcaactttga aattaactac agacctggat cagcaaatca catagctgat 1740
gccttatcca gaattgttga cgaaacagaa ccaattccaa aagattcaga agacaatagt 1800
atcaactttg ttaatcaaat ctcgataacc gatgatttta aaaaccaagt ggttacagaa 1860
tatacgaatg atacaaaatt gttgaattta ctaaacaatg aagacaaacg agtggaagag 1920
aatatccaac tcaaagatgg cttactaatt aacagtaaag accaaatctt attacctaat 1980
gatactcagc tgactaggac aattattaaa aagtatcatg aagaaggtaa attgattcat 2040
ccaggcattg aacttcttac aaacattata ttacgtagat ttacgtggaa aggaataaga 2100
aaacaaatac aagaatatgt acagaactgc catacatgtc aaataaacaa atctaggaat 2160
cataaacctt atggaccttt acaaccaatt cccccatcag aaagaccttg ggaatcttta 2220
tcaatggatt ttattacagc tttaccagaa tcatctggtt ataatgcact tttcgtggta 2280
gttgaccgat tttcaaaaat ggcaatctta gtaccttgta cgaaatccat tacagcagag 2340
caaacagctc gaatgtttga tcaacgagtt attgcttatt tcggcaatcc aaaagaaatc 2400
attgcagata atgatcatat ttttacttcc caaacgtgga aagatttcgc acataaatat 2460
aatttcgtta tgaaattttc gttaccatac agaccacaaa ctgatggaca aactgagcgt 2520
acaaaccaaa ctgtggagaa attactaaga tgtgtatgta gcacacatcc aaatacatgg 2580
gtagatcata tatccctagt gcaacaatct tacaacaatg cgatacattc agcaactcaa 2640
atgacacctt ttgagatagt acatcgctat tcaccagctt tatcaccttt agagttacct 2700
agctttagtg acaaaactga cgaaaactct caggaaacga tccaagtatt tcaaacagtt 2760

aaagaacact tgaatacaaa caacataaag atgaaaaagt atttcgatat gaaaatacaa 2820
gaaattgaag aatttcaacc tggagaccta gttatggtca aaagaacgaa aacagcattc 2880
ttatacacca ataacagaca aacagagagc caaatcatga gtgaactccc attcacaatt 2940
gcttcaaaga gaataaaata cctaggaatt caactgacaa gggaagtgaa ggacctcttc 3000
aaggagaact acaaa 3015

<210> 46
<211> 585
<212> DNA
<213> Homo sapiens

<400> 46
ggccccagac tggcccacgg caccaccact ctggccttcc gcttccgtca tggagtcatt 60
gctgcagctg acacgcgttc ctcctgtggc agctatgtgg cgtgtccagc ctcatgcaag 120
gtcatccctg tgcaccagca cctcctgggt accacctctg gcacctctgc cgactgtgct 180
acctggtatc gggtattaca gcgggagctg cggcttcggg aactgaggga gggtcagctg 240
cccagtgtgg ccagtgctgc caagctcttg tcagccatga tgtctcaata ccggggactg 300
gatctctgtg tggccactgc cctctgcggc tgggaccgct ctggccctga gctcttctac 360
gtctatagcg acggcacccg cctgcagggg gacatcttct ctgtgggctc tggatctccc 420
tatgcctacg gcgtgctaga ccgtggctat cgctacgaca tgagcaccca ggaagcctac 480
gccctggctc gctgcgccgt ggcccacgcc acccaccgtg atgcctattc agggggctct 540
gtagaccttt tccacgtgcg ggagagtgga tgggagcatg tgtca 585

<210> 47
<211> 594
<212> DNA
<213> Homo sapiens

<400> 47
tctattatgt cctataacgg aggagccatc atggccatga aggggaagaa ccgtgtggcc 60
atcgctgcag acaggcactt cgggatccag gcccagatgg tgaccacgga cttccaggag 120
atctttccca tgggtggttg gttgtacatc ggtctggccg ggcttgccac tgacgtccag 180
agagttgccc agtgcctcaa gttccagctg aacctatatg agttgaagga aggtcagcag 240
atcaaacctt ataccttcac gagcatggtg gccaacttct tgtatgagaa acattttggc 300
ccctactaca ctgatccagt cattgctggt ttggacctga agacctttaa gcccttcagt 360
tgctctctag acctcatcgg cttccccatg gtgactgatg actttgtggt caatggcagc 420
tatgccgaac aaatgtacgg aatgtgtgag tccctctggg aacccaacat ggatccagaa 480
cacccgtttg aaaccatctc cccagccatg ctgaatgctg tggactgggg tgcagggtca 540
ggcatgggag tcatcatcca catcaccaag aaggacaaaa tcaccaccag gaca 594

<210> 48
<211> 361
<212> PRT
<213> Homo sapiens

<400> 48
Leu Val Asp Glu Gln Pro Leu Glu Asn Tyr Leu Asp Met Glu Tyr Phe
1 5 10 15
Gly Thr Ile Gly Ile Gly Thr Pro Ala Gln Asp Phe Thr Val Val Phe
20 25 30
Asp Thr Gly Ser Ser Asn Leu Trp Val Pro Ser Val Tyr Cys Ser Ser
35 40 45
Leu Ala Cys Thr Asn His Asn Arg Phe Asn Pro Glu Asp Ser Ser Thr
50 55 60
Tyr Gln Ser Thr Ser Glu Thr Val Ser Ile Thr Tyr Gly Thr Gly Ser
65 70 75 80
Met Thr Gly Ile Leu Gly Tyr Asp Thr Val Gln Val Gly Gly Ile Ser
85 90 95
Asp Thr Asn Gln Ile Phe Gly Leu Ser Glu Thr Glu Pro Gly Ser Phe
100 105 110
Leu Tyr Tyr Ala Pro Phe Asp Gly Ile Leu Gly Leu Ala Tyr Pro Ser
115 120 125

```
Ile Ser Ser Ser Gly Ala Thr Pro Val Phe Asp Asn Ile Trp Asn Gln
    130                 135                 140
Gly Leu Val Ser Gln Asp Leu Phe Ser Val Tyr Leu Ser Ala Asp Asp
145                 150                 155                 160
Lys Ser Gly Ser Val Val Ile Phe Gly Gly Ile Asp Ser Ser Tyr Tyr
                165                 170                 175
Thr Gly Ser Leu Asn Trp Val Pro Val Thr Val Glu Gly Tyr Trp Gln
            180                 185                 190
Ile Thr Val Asp Ser Ile Thr Met Asn Gly Glu Thr Ile Ala Cys Ala
        195                 200                 205
Glu Gly Cys Gln Ala Ile Val Asp Thr Gly Thr Ser Leu Leu Thr Gly
    210                 215                 220
Pro Thr Ser Pro Ile Ala Asn Ile Gln Ser Asp Ile Gly Ala Ser Glu
225                 230                 235                 240
Asn Ser Asp Gly Asp Val Ser Pro Ala Pro Thr Ala Leu Phe Tyr Thr
                245                 250                 255
Gln Val Val Gly Val Pro Gly Arg Ser Asp Glu Asn Pro Ser Asn Phe
            260                 265                 270
Ser His Pro His Ser Phe Gln Met Val Val Ser Cys Ser Ala Ile Ser
        275                 280                 285
Ser Leu Pro Asp Ile Val Phe Thr Ile Asn Gly Val Gln Tyr Pro Val
    290                 295                 300
Pro Pro Ser Ala Tyr Ile Leu Gln Ser Glu Gly Ser Cys Ile Ser Gly
305                 310                 315                 320
Phe Gln Gly Met Asn Val Pro Thr Glu Ser Gly Glu Leu Trp Ile Leu
                325                 330                 335
Gly Asp Val Phe Ile Arg Gln Tyr Phe Thr Val Phe Asp Arg Ala Asn
            340                 345                 350
Asn Gln Val Gly Leu Ala Pro Val Ala
        355                 360
```

<210> 49
<211> 361
<212> PRT
<213> Homo sapiens

<400> 49

```
Leu Val Asp Glu Gln Pro Leu Glu Asn Tyr Leu Asp Met Glu Tyr Phe
 1               5                  10                  15
Gly Thr Ile Gly Ile Gly Thr Pro Ala Gln Asp Phe Thr Val Leu Phe
            20                  25                  30
Asp Thr Gly Ser Ser Asn Leu Trp Val Pro Ser Val Tyr Cys Ser Ser
        35                  40                  45
Leu Ala Cys Thr Asn His Asn Arg Phe Asn Pro Glu Asp Ser Ser Thr
    50                  55                  60
Tyr Gln Ser Thr Ser Glu Thr Val Ser Ile Thr Tyr Gly Thr Gly Ser
65                  70                  75                  80
Met Thr Gly Ile Leu Gly Tyr Asp Thr Val Gln Val Gly Gly Ile Ser
                85                  90                  95
Asp Thr Asn Gln Ile Phe Gly Leu Ser Glu Thr Glu Pro Gly Ser Phe
            100                 105                 110
Leu Tyr Tyr Ala Pro Phe Asp Gly Ile Leu Gly Leu Ala Tyr Pro Ser
        115                 120                 125
Ile Ser Ser Ser Gly Ala Thr Pro Val Phe Asp Asn Ile Trp Asn Gln
    130                 135                 140
Gly Leu Val Ser Gln Asp Leu Phe Ser Val Tyr Leu Ser Ala Asp Asp
145                 150                 155                 160
Lys Ser Gly Ser Val Val Ile Phe Gly Gly Ile Asp Ser Ser Tyr Tyr
                165                 170                 175
Thr Gly Ser Leu Asn Trp Val Pro Val Thr Val Glu Gly Tyr Trp Gln
            180                 185                 190
```

```
Ile Thr Val Asp Ser Ile Thr Met Asn Gly Glu Thr Ile Ala Cys Ala
        195                 200                 205
Glu Gly Cys Gln Ala Ile Val Asp Thr Gly Thr Ser Leu Leu Thr Gly
    210                 215                 220
Pro Thr Ser Pro Ile Ala Asn Ile Gln Ser Asp Ile Gly Ala Ser Glu
225                 230                 235                 240
Asn Ser Asp Gly Asp Val Ser Pro Ala Pro Thr Ala Leu Phe Tyr Thr
                245                 250                 255
Gln Val Val Gly Val Pro Gly Arg Ser Asp Glu Asn Pro Ser Asn Phe
            260                 265                 270
Ser His Pro His Ser Phe Gln Met Val Val Ser Cys Ser Ala Ile Ser
        275                 280                 285
Ser Leu Pro Asp Ile Val Phe Thr Ile Asn Gly Val Gln Tyr Pro Val
    290                 295                 300
Pro Pro Ser Ala Tyr Ile Leu Gln Ser Glu Gly Ser Cys Ile Ser Gly
305                 310                 315                 320
Phe Gln Gly Met Asn Val Pro Thr Glu Ser Gly Glu Leu Trp Ile Leu
                325                 330                 335
Gly Asp Val Phe Ile Arg Gln Tyr Phe Thr Val Phe Asp Arg Ala Asn
            340                 345                 350
Asn Gln Val Gly Leu Ala Pro Val Ala
        355                 360
```

<210> 50
<211> 305
<212> PRT
<213> Homo sapiens

```
<400> 50
Asp Lys Ala Trp Leu Lys Arg Gly Asn Lys Gln Phe Asn Glu Gly Lys
 1               5                  10                  15
Glu Ser Asp Arg Cys Leu Ile Phe Lys Cys Lys Asn Lys Asp Val Lys
            20                  25                  30
Met Ile Glu Gln His Asn Gln Glu Tyr Ser Gln Gly Lys His Ser Phe
        35                  40                  45
Thr Met Ala Met Asn Ala Phe Gly Asp Met Thr Asn Glu Glu Phe Arg
    50                  55                  60
Gln Val Met Asn Gly Phe Gln Tyr Gln Lys His Arg Lys Gly Lys Gln
65                  70                  75                  80
Phe Gln Glu Arg Leu Leu Leu Glu Ile Pro Thr Ser Val Asp Trp Arg
                85                  90                  95
Glu Lys Gly Tyr Met Thr Pro Val Lys Asp Gln Gly Gln Cys Gly Ser
            100                 105                 110
Cys Trp Ala Phe Ser Ala Thr Gly Ala Leu Glu Gly Gln Met Phe Trp
        115                 120                 125
Lys Thr Gly Lys Leu Ile Ser Leu Asn Glu Gln Asn Leu Val Asp Cys
    130                 135                 140
Ser Gly Pro Gln Gly Asn Glu Gly Cys Asn Gly Asp Phe Met Asp Asn
145                 150                 155                 160
Pro Phe Arg Tyr Val Gln Glu Asn Gly Gly Leu Asp Ser Glu Ala Ser
                165                 170                 175
Tyr Pro Tyr Glu Gly Lys Val Lys Thr Cys Arg Tyr Asn Pro Lys Tyr
            180                 185                 190
Ser Ala Ala Asn Asp Thr Gly Phe Val Asp Ile Pro Ser Arg Glu Lys
        195                 200                 205
Asp Leu Ala Lys Ala Val Ala Thr Val Gly Pro Ile Ser Val Ala Val
    210                 215                 220
Gly Ala Ser His Val Phe Phe Gln Phe Tyr Lys Lys Gly Ile Tyr Phe
225                 230                 235                 240
Glu Pro Arg Cys Asp Pro Glu Gly Leu Asp His Ala Met Leu Val Val
                245                 250                 255
```

```
Gly Tyr Ser Tyr Glu Gly Ala Asp Ser Asp Asn Asn Lys Tyr Trp Leu
            260                 265                 270
Val Lys Asn Arg Glu Ser Phe Phe Ala Ile Ser Ile Tyr Ile Pro Gln
        275                 280                 285
Ala Leu Asn His Ile Ser Glu Ile Pro Glu Val Phe Phe Pro Leu Arg
    290                 295                 300
Gln
305
```

<210> 51
<211> 190
<212> PRT
<213> Homo sapiens

<400> 51

```
Ser Asn Val Lys Pro Arg Lys Lys Asn Pro Lys Gly Phe Thr Val Asn
 1               5                  10                  15
Val Leu Ala Asn Ile Gly Val Ile Asn His Trp Lys Leu Pro Lys Leu
            20                  25                  30
Thr His Lys Gly Glu Lys Asn Ser Gly Lys Glu Gly Glu Lys Pro His
        35                  40                  45
Lys Thr Asn Ser Ala Ser Asn Ser Cys Gly Thr Ile Ser Asn Leu Gly
    50                  55                  60
Ser Ile Glu Asn Gly Thr Asn Glu Gly Lys Tyr Lys Ile Asn Phe Ile
65                  70                  75                  80
Leu Lys Val Ile Leu Lys Asn Ala Asn Thr Val Arg Thr Glu Lys Arg
                85                  90                  95
Asn Arg Gln Ile His Asn Tyr Asn Cys Leu His Cys Thr Val Glu Thr
            100                 105                 110
Ser Asp Tyr Gln Val Gln Ala Pro Ser Ile Asp Glu Lys Val Asp Leu
        115                 120                 125
His Phe Ile Ala Leu Val His Val Asp Gly His Leu Tyr Glu Leu Asp
    130                 135                 140
Gly Arg Lys Pro Phe Pro Ile Asn His Gly Glu Thr Ser Asp Glu Thr
145                 150                 155                 160
Leu Leu Glu Asp Ala Ile Glu Val Cys Lys Lys Phe Met Glu Arg Asp
                165                 170                 175
Pro Asp Glu Leu Arg Phe Asn Ala Ile Ala Leu Ser Ala Ala
            180                 185                 190
```

<210> 52
<211> 204
<212> PRT
<213> Homo sapiens

<400> 52

```
Ala Cys Leu Val Ser Gly Phe Gly Pro Phe Arg Gln His Leu Val Asn
 1               5                  10                  15
Ser Ser Trp Glu Ala Val Lys Glu Leu Ser Lys Leu Gly Leu Gly Asn
            20                  25                  30
Glu Thr Val Val Gln Leu Arg Thr Leu Glu Leu Pro Val Asp Tyr Arg
        35                  40                  45
Glu Ala Lys Arg Arg Val Thr Gly Ile Trp Glu Asp His Gln Pro Gln
    50                  55                  60
Leu Val Val His Val Gly Met Asp Thr Ala Ala Lys Ala Ile Ile Leu
65                  70                  75                  80
Glu Gln Ser Gly Lys Asn Gln Gly Tyr Arg Asp Ala Asp Ile Arg Ser
                85                  90                  95
Phe Trp Pro Glu Gly Gly Val Cys Leu Pro Gly Ser Pro Asp Val Leu
            100                 105                 110
```

```
Glu Ser Gly Val Cys Met Lys Ala Val Cys Lys Arg Val Ala Val Glu
        115                 120                 125
Gly Val Asp Val Ile Phe Ser Arg Asp Ala Gly Arg Tyr Val Cys Asp
    130                 135                 140
Tyr Thr Tyr Tyr Leu Ser Leu His His Gly Lys Gly Cys Ala Ala Leu
145                 150                 155                 160
Ile His Val Pro Pro Leu Ser Arg Gly Leu Pro Ala Ser Leu Leu Gly
                165                 170                 175
Arg Ala Leu Arg Val Ile Ile Gln Glu Met Leu Glu Glu Ala Gly Glu
            180                 185                 190
Lys Gln Lys Glu Val Thr Ala Ser Gly Thr Ser His
        195                 200
```

<210> 53
<211> 358
<212> PRT
<213> Homo sapiens

<400> 53

```
Arg Lys Lys Ser Val Tyr Thr Val Gly Leu Arg Gly Leu Ile Asn Leu
 1               5                  10                  15
Gly Asn Thr Cys Phe Met Asn Cys Ile Val Gln Ala Leu Thr His Ile
            20                  25                  30
Pro Leu Leu Lys Asp Phe Phe Leu Ser Asp Lys His Lys Cys Ile Met
        35                  40                  45
Thr Ser Pro Ser Leu Cys Leu Val Cys Glu Met Ser Ser Leu Phe His
    50                  55                  60
Ala Met Tyr Ser Gly Ser Arg Thr Pro His Ile Pro Tyr Lys Leu Leu
65                  70                  75                  80
His Leu Ile Trp Ile His Ala Glu His Leu Ala Gly Tyr Arg Gln Gln
                85                  90                  95
Asp Ala His Glu Phe Leu Ile Ala Ile Leu Asp Val Leu His Arg His
            100                 105                 110
Ser Lys Asp Asp Ser Gly Gly Gln Glu Ala Asn Asn Pro Asn Cys Cys
        115                 120                 125
Asn Cys Ile Ile Asp Gln Ile Phe Thr Gly Gly Leu Gln Ser Asp Val
    130                 135                 140
Thr Cys Gln Ala Cys His Ser Val Ser Thr Thr Ile Asp Pro Cys Trp
145                 150                 155                 160
Asp Ile Ser Leu Asp Leu Pro Gly Ser Cys Ala Thr Phe Asp Ser Gln
                165                 170                 175
Asn Pro Glu Arg Ala Asp Ser Thr Val Ser Arg Asp Asp His Ile Pro
            180                 185                 190
Gly Ile Pro Ser Leu Thr Asp Cys Leu Gln Trp Phe Thr Arg Pro Glu
        195                 200                 205
His Leu Gly Ser Ser Ala Lys Ile Lys Cys Asn Ser Cys Gln Ser Tyr
    210                 215                 220
Gln Glu Ser Thr Lys Gln Leu Thr Met Lys Lys Leu Pro Ile Val Ala
225                 230                 235                 240
Cys Phe His Leu Lys Arg Phe Glu His Val Gly Lys Gln Arg Arg Lys
                245                 250                 255
Ile Asn Thr Phe Ile Ser Phe Pro Leu Glu Leu Asp Met Thr Pro Phe
            260                 265                 270
Leu Ala Ser Thr Lys Glu Ser Arg Met Lys Glu Gly Gln Pro Pro Thr
        275                 280                 285
Asp Cys Val Pro Asn Glu Asn Lys Tyr Ser Leu Phe Ala Val Ile Asn
    290                 295                 300
His His Gly Thr Leu Glu Ser Gly His Tyr Thr Ser Phe Ile Arg Gln
305                 310                 315                 320
Gln Lys Asp Gln Trp Phe Ser Cys Asp Asp Ala Ile Ile Thr Lys Ala
                325                 330                 335
```

Thr Ile Glu Asp Leu Leu Tyr Ser Glu Gly Tyr Leu Leu Phe Tyr His
340 345 350
Lys Gln Gly Leu Glu Lys
355

<210> 54
<211> 358
<212> PRT
<213> Homo sapiens

<400> 54
Arg Lys Lys Ser Val Tyr Thr Val Gly Leu Arg Gly Leu Ile Asn Leu
1 5 10 15
Gly Asn Thr Cys Phe Met Asn Cys Ile Val Gln Ala Leu Thr His Ile
20 25 30
Pro Leu Leu Lys Asp Phe Phe Leu Ser Asp Lys His Lys Cys Ile Met
35 40 45
Thr Ser Pro Ser Leu Cys Leu Val Cys Glu Met Ser Ser Leu Phe His
50 55 60
Ala Met Tyr Ser Gly Ser Arg Thr Pro His Ile Pro Tyr Lys Leu Leu
65 70 75 80
His Leu Ile Trp Ile His Ala Glu His Leu Ala Gly Tyr Arg Gln Gln
85 90 95
Asp Ala His Glu Phe Leu Ile Ala Ile Leu Asp Val Leu His Arg His
100 105 110
Ser Lys Asp Asp Ser Gly Gly Gln Glu Ala Asn Asn Pro Asn Cys Cys
115 120 125
Asn Cys Ile Ile Asp Gln Ile Phe Thr Gly Gly Leu Gln Ser Asp Val
130 135 140
Thr Cys Gln Ala Cys His Ser Val Ser Thr Thr Ile Asp Pro Cys Trp
145 150 155 160
Asp Ile Ser Leu Asp Leu Pro Gly Ser Cys Ala Thr Phe Asp Ser Gln
165 170 175
Asn Pro Glu Arg Ala Asp Ser Thr Val Ser Arg Asp Asp His Ile Pro
180 185 190
Gly Ile Pro Ser Leu Thr Asp Cys Leu Gln Trp Phe Thr Arg Pro Glu
195 200 205
His Leu Gly Ser Ser Ala Lys Ile Lys Cys Asn Ser Cys Gln Ser Tyr
210 215 220
Gln Glu Ser Thr Lys Gln Leu Thr Met Lys Lys Leu Pro Ile Val Ala
225 230 235 240
Cys Phe His Leu Lys Arg Phe Glu His Val Gly Lys Gln Arg Arg Lys
245 250 255
Ile Asn Thr Phe Ile Ser Phe Pro Leu Glu Leu Asp Met Thr Pro Phe
260 265 270
Leu Ala Ser Thr Lys Glu Ser Arg Met Lys Glu Gly Gln Pro Pro Thr
275 280 285
Asp Cys Val Pro Asn Glu Asn Lys Tyr Ser Leu Phe Ala Val Ile Asn
290 295 300
His His Gly Thr Leu Glu Ser Gly His Tyr Thr Ser Phe Ile Arg Gln
305 310 315 320
Gln Lys Asp Gln Trp Phe Ser Cys Asp Asp Ala Ile Ile Thr Lys Ala
325 330 335
Thr Ile Glu Asp Leu Leu Tyr Ser Glu Gly Tyr Leu Leu Phe Tyr His
340 345 350
Lys Gln Gly Leu Glu Lys
355

<210> 55
<211> 357

<212> PRT
<213> Homo sapiens

<400> 55
Arg Ile Thr Ser Ser Phe Thr Ile Gly Leu Arg Gly Leu Ile Asn Leu
1 5 10 15
Gly Asn Thr Cys Phe Met Asn Cys Ile Val Gln Ala Leu Thr His Thr
20 25 30
Pro Ile Leu Arg Asp Phe Phe Leu Ser Asp Arg His Arg Cys Glu Met
35 40 45
Pro Ser Pro Glu Leu Cys Leu Val Cys Glu Met Ser Ser Leu Phe Arg
50 55 60
Glu Leu Tyr Ser Gly Asn Pro Ser Pro His Val Pro Tyr Lys Leu Leu
65 70 75 80
His Leu Val Trp Ile His Ala Arg His Leu Ala Gly Tyr Arg Gln Gln
85 90 95
Asp Ala His Glu Phe Leu Ile Ala Ala Leu Asp Val Leu His Arg His
100 105 110
Cys Lys Gly Asp Asp Val Gly Lys Ala Ala Asn Asn Pro Asn His Cys
115 120 125
Asn Cys Ile Ile Asp Gln Ile Phe Thr Gly Gly Leu Gln Ser Asp Val
130 135 140
Thr Cys Gln Ala Cys His Gly Val Ser Thr Thr Ile Asp Pro Cys Trp
145 150 155 160
Asp Ile Ser Leu Asp Leu Pro Gly Ser Cys Thr Ser Phe Trp Pro Met
165 170 175
Ser Pro Gly Arg Glu Ser Ser Val Asn Gly Glu Ser His Ile Pro Gly
180 185 190
Ile Thr Thr Leu Thr Asp Cys Leu Arg Arg Phe Thr Arg Pro Glu His
195 200 205
Leu Gly Ser Ser Ala Lys Ile Lys Cys Gly Ser Cys Gln Ser Tyr Gln
210 215 220
Glu Ser Thr Lys Gln Leu Thr Met Asn Lys Leu Pro Val Val Ala Cys
225 230 235 240
Phe His Phe Lys Arg Phe Glu His Ser Ala Lys Gln Arg Arg Lys Ile
245 250 255
Thr Thr Tyr Ile Ser Phe Pro Leu Glu Leu Asp Met Thr Pro Phe Met
260 265 270
Ala Ser Ser Lys Glu Ser Arg Met Asn Gly Gln Leu Gln Leu Pro Thr
275 280 285
Asn Ser Gly Asn Asn Glu Asn Lys Tyr Ser Leu Phe Ala Val Val Asn
290 295 300
His Gln Gly Thr Leu Glu Ser Gly His Tyr Thr Ser Phe Ile Arg His
305 310 315 320
His Lys Asp Gln Trp Phe Lys Cys Asp Asp Ala Val Ile Thr Lys Ala
325 330 335
Ser Ile Lys Asp Val Leu Asp Ser Glu Gly Tyr Leu Leu Phe Tyr His
340 345 350
Lys Gln Val Leu Glu
355

<210> 56
<211> 308
<212> PRT
<213> Homo sapiens

<400> 56
Leu Ser Ser Arg Arg Pro Ala Ala Val Gly Ala Gly Leu Gln Asn Met
1 5 10 15
Gly Asn Thr Cys Tyr Glu Asn Ala Ser Leu Gln Cys Leu Thr Tyr Thr
20 25 30

```
Pro Pro Leu Ala Asn Tyr Met Leu Ser Arg Glu His Ser Gln Thr Cys
        35                  40                  45
Gln Arg Pro Lys Cys Cys Met Leu Cys Thr Met Gln Ala His Ile Thr
    50                  55                  60
Trp Ala Leu His Ser Pro Gly His Val Ile Gln Pro Ser Gln Ala Leu
65                  70                  75                  80
Ala Ala Gly Phe His Arg Gly Lys Gln Glu Asp Ala His Glu Phe Leu
                85                  90                  95
Met Phe Thr Val Asp Ala Met Lys Lys Ala Cys Leu Pro Gly His Lys
            100                 105                 110
Gln Val Asp His His Ser Lys Asp Thr Thr Leu Ile His Gln Ile Phe
        115                 120                 125
Gly Gly Cys Trp Arg Ser Gln Ile Lys Cys Leu His Cys His Gly Ile
    130                 135                 140
Ser Asp Thr Phe Asp Pro Tyr Leu Asp Ile Ala Leu Asp Ile Gln Ala
145                 150                 155                 160
Ala Gln Ser Val Lys Gln Ala Leu Glu Gln Leu Val Lys Pro Glu Glu
                165                 170                 175
Leu Asn Gly Glu Asn Ala Tyr His Cys Gly Leu Cys Leu Gln Arg Ala
            180                 185                 190
Pro Ala Ser Lys Thr Leu Thr Leu His Thr Ser Ala Lys Val Leu Ile
        195                 200                 205
Leu Val Leu Lys Arg Phe Ser Asp Val Thr Gly Asn Lys Leu Ala Lys
    210                 215                 220
Asn Val Gln Tyr Pro Glu Cys Leu Asp Met Gln Pro Tyr Met Ser Gln
225                 230                 235                 240
Gln Asn Thr Gly Pro Leu Val Tyr Val Leu Tyr Ala Val Leu Val His
                245                 250                 255
Ala Gly Trp Ser Cys His Asp Gly His Tyr Phe Ser Tyr Val Lys Ala
            260                 265                 270
Gln Glu Gly Gln Trp Tyr Lys Met Asp Asp Ala Lys Val Thr Ala Cys
        275                 280                 285
Ser Ile Thr Ser Val Leu Ser Gln Gln Ala Tyr Val Leu Phe Tyr Ile
    290                 295                 300
Gln Lys Ser Glu
305
```

<210> 57
<211> 337
<212> PRT
<213> Homo sapiens

<400> 57

```
Cys Glu Lys Arg Glu Asn Leu Leu Pro Phe Val Gly Leu Asn Asn Leu
 1               5                  10                  15
Gly Asn Thr Cys Tyr Leu Asn Ser Ile Leu Gln Val Leu Tyr Phe Cys
            20                  25                  30
Pro Gly Phe Met Tyr Cys Ile Phe Lys Thr Arg Ile Asp Glu Met Glu
        35                  40                  45
Ile Phe Ile Tyr Arg Glu Leu Asn Pro Met Tyr Glu Gly Tyr Leu Gln
    50                  55                  60
His Asp Ala Gln Glu Val Leu Gln Cys Ile Leu Gly Asn Ile Gln Glu
65                  70                  75                  80
Thr Cys Gln Leu Leu Lys Lys Glu Glu Val Lys Asn Val Ala Glu Leu
                85                  90                  95
Pro Thr Lys Val Glu Glu Ile Pro His Pro Lys Glu Glu Met Asn Gly
            100                 105                 110
Glu Glu Gln Ile Gly Phe Glu Leu Val Glu Lys Leu Phe Gln Gly Gln
        115                 120                 125
Leu Val Leu Arg Thr Arg Cys Leu Glu Cys Glu Ser Leu Thr Glu Arg
    130                 135                 140
```

```
Arg Glu Asp Phe Gln Asp Ile Ser Val Pro Val Gln Glu Asp Glu Leu
145                 150                 155                 160
Ser Lys Val Glu Glu Ser Ser Glu Lys Met Lys Thr Leu Arg Trp Ala
                165                 170                 175
Ile Ser Gln Phe Ala Ser Val Glu Arg Ile Val Gly Glu Asp Lys Tyr
            180                 185                 190
Phe Cys Glu Asn Cys His His Tyr Thr Glu Ala Glu Arg Ser Leu Leu
        195                 200                 205
Phe Asp Lys Met Pro Glu Val Ile Thr Ile His Leu Lys Cys Phe Ala
    210                 215                 220
Ala Ser Gly Asn Leu Phe Phe Phe Ser Ser Gln Arg Phe Asp Cys Tyr
225                 230                 235                 240
Gly Gly Gly Leu Ser Lys Ile Asn Thr Pro Leu Leu Thr Pro Leu Lys
                245                 250                 255
Leu Ser Leu Glu Glu Trp Ser Thr Lys Pro Thr Asn Asp Ser Tyr Gly
            260                 265                 270
Leu Phe Ala Val Val Met His Ser Gly Ile Thr Ile Ser Ser Gly His
        275                 280                 285
Tyr Thr Ala Ser Val Lys Glu Tyr Glu Gly Lys Trp Leu Leu Phe Asp
    290                 295                 300
Asp Ser Glu Val Lys Val Thr Glu Glu Lys Asp Phe Leu Asn Ser Leu
305                 310                 315                 320
Ser Pro Ser Thr Ser Pro Thr Ser Thr Pro Tyr Leu Leu Phe Tyr Lys
                325                 330                 335
Lys
```

<210> 58
<211> 308
<212> PRT
<213> Homo sapiens

```
<400> 58
Leu Ser Ser Arg Arg Pro Ala Ala Val Gly Ala Gly Leu Gln Asn Met
 1               5                  10                  15
Gly Asn Thr Cys Tyr Glu Asn Ala Ser Leu Gln Cys Leu Thr Tyr Thr
            20                  25                  30
Leu Pro Leu Ala Asn Tyr Met Leu Ser Arg Glu His Ser Gln Thr Cys
        35                  40                  45
Gln Arg Pro Lys Cys Cys Met Leu Cys Thr Met Gln Ala His Ile Thr
    50                  55                  60
Trp Ala Leu His Ser Pro Gly His Val Ile Gln Pro Ser Gln Ala Leu
65                  70                  75                  80
Ala Ala Gly Phe His Arg Gly Lys Gln Glu Asp Val His Glu Phe Leu
                85                  90                  95
Met Phe Thr Val Asp Ala Met Lys Lys Ala Cys Leu Pro Gly His Lys
            100                 105                 110
Gln Val Asp His His Cys Lys Asp Thr Thr Leu Ile His Gln Ile Phe
        115                 120                 125
Gly Gly Cys Trp Arg Ser Gln Ile Lys Cys Leu His Cys His Gly Ile
    130                 135                 140
Ser Asp Thr Phe Asp Pro Tyr Leu Asp Ile Ala Leu Asp Ile Gln Ala
145                 150                 155                 160
Ala Gln Ser Val Lys Gln Ala Leu Glu Gln Leu Val Lys Pro Glu Glu
                165                 170                 175
Leu Asn Gly Glu Asn Ala Tyr His Cys Gly Leu Cys Leu Gln Arg Ala
            180                 185                 190
Pro Ala Ser Asn Thr Leu Thr Leu His Thr Ser Ala Lys Val Leu Ile
        195                 200                 205
Leu Val Leu Lys Arg Phe Ser Asp Val Ala Gly Asn Lys Leu Ala Lys
    210                 215                 220
```

```
Asn Val Gln Tyr Pro Glu Cys Leu Asp Met Gln Pro Tyr Met Ser Gln
225                 230                 235                 240
Gln Asn Thr Gly Pro Leu Val Tyr Val Leu Tyr Ala Val Leu Val His
            245                 250                 255
Ala Gly Trp Ser Cys His Asp Gly His Tyr Phe Ser Tyr Val Lys Ala
        260                 265                 270
Gln Glu Val Gln Trp Tyr Lys Met Asp Asp Ala Glu Val Thr Val Cys
    275                 280                 285
Ser Ile Ile Ser Val Leu Ser Gln Gln Ala Tyr Val Leu Phe Tyr Ile
290                 295                 300
Gln Lys Ser Glu
305
```

<210> 59
<211> 232
<212> PRT
<213> Homo sapiens

<400> 59

```
Tyr Asp Arg Lys Arg Gln Asp Lys Ala Pro Val Gly Leu Lys Asn Val
 1               5                  10                  15
Gly Asn Thr Cys Trp Phe Ser Ala Val Ile Gln Ser Leu Phe Asn Leu
            20                  25                  30
Leu Glu Phe Arg Arg Leu Val Leu Asn Tyr Lys Pro Pro Ser Asn Ala
        35                  40                  45
Gln Asp Leu Pro Arg Asn Gln Lys Ala Phe Phe Phe Ser Gln Gln Asp
    50                  55                  60
Val Ser Glu Phe Thr His Lys Leu Leu Asp Trp Leu Glu Asp Ala Phe
65                  70                  75                  80
Gln Met Lys Ala Glu Glu Glu Thr Val Gly Lys Asp Val Glu Lys Leu
                85                  90                  95
Lys Pro Leu Cys Ser Val Gly Glu Asp Met Lys Trp Tyr Ser His Cys
            100                 105                 110
Gly Lys His Phe Cys Tyr Cys Phe Ile Ser Phe Gln His Trp Phe Thr
        115                 120                 125
Glu Leu Pro Pro Val Leu Thr Phe Glu Leu Ser Arg Phe Glu Phe Asn
    130                 135                 140
Gln Ala Leu Gly Arg Pro Glu Lys Ile His Asn Lys Leu Glu Phe Pro
145                 150                 155                 160
Gln Val Pro Tyr Arg Leu His Ala Val Leu Val His Glu Gly Gln Ala
                165                 170                 175
Asn Ala Gly His Tyr Trp Ala Tyr Ile Phe Asp His Arg Glu Ser Arg
            180                 185                 190
Trp Met Lys Tyr Asn Asp Ile Ala Val Thr Lys Ser Ser Trp Glu Glu
        195                 200                 205
Leu Val Arg Asp Ser Phe Gly Gly Tyr Arg Asn Ala Ser Ala Tyr Cys
    210                 215                 220
Leu Met Tyr Ile Asn Asp Lys Ala
225                 230
```

<210> 60
<211> 228
<212> PRT
<213> Homo sapiens

<400> 60

```
Asn Asp Trp Arg Arg Val Asp Gly Trp Pro Val Gly Leu Lys Asn Val
 1               5                  10                  15
Gly Asn Thr Cys Trp Phe Ser Ala Val Ile Gln Ser Leu Phe Gln Leu
            20                  25                  30
```

```
Pro Glu Phe Arg Arg Leu Val Leu Ser Tyr Ser Leu Pro Gln Asn Val
        35                  40                  45
Leu Glu Asn Cys Arg Ser His Thr Glu Gln Gln Gln Asp Val Ser Glu
    50                  55                  60
Phe Thr His Lys Leu Leu Asp Trp Leu Glu Asp Ala Phe Gln Leu Ala
65                  70                  75                  80
Val Asn Thr Phe Gly Gln Tyr Pro Leu Gln Val Asn Gly Tyr Arg Asn
                85                  90                  95
Leu Asp Glu Cys Leu Glu Gly Ala Met Val Glu Gly Asp Val Glu Leu
            100                 105                 110
Leu Pro Ser Asp His Ser Val Lys Tyr Gly Gln Glu Arg Trp Phe Thr
        115                 120                 125
Lys Leu Pro Pro Val Leu Thr Phe Glu Leu Ser Arg Phe Glu Phe Asn
    130                 135                 140
His Ser Trp Gly Arg Asp Lys Lys Asp Ser Lys Ala Leu His Thr Val
145                 150                 155                 160
Pro Tyr Arg Leu His Ala Val Leu Val His Glu Gly Gln Ala Asn Ala
                165                 170                 175
Gly His Tyr Trp Ala Tyr Ile Tyr Asn Gln Pro Arg Gln Ser Trp Leu
            180                 185                 190
Lys Tyr Asn Asp Ile Ser Val Thr Glu Ser Ser Trp Glu Glu Val Glu
        195                 200                 205
Arg Asp Ser Tyr Gly Gly Leu Arg Asn Val Ser Ala Tyr Cys Leu Met
    210                 215                 220
Tyr Ile Asn Asp
225
```

<210> 61
<211> 256
<212> PRT
<213> Homo sapiens

<400> 61

```
Ser Gly Ser Ser Pro Ser Ser Ser Trp Pro Ser Gly Leu Arg Ser Ser
 1               5                  10                  15
Cys Pro Ile Phe Gln Cys Leu Phe Met Leu His Leu Leu Ser Arg Ser
            20                  25                  30
Gln Ser Phe Leu Trp Pro Arg Val Arg Met Arg Arg Gln His Gly Ala
        35                  40                  45
Leu Glu Phe His Arg Val Leu Phe Gly Ser Leu Gln Glu Glu Arg Ala
    50                  55                  60
Gln Asp Ala Asp Ser Val Trp Gln Gln Gln Gln Ala His Gln Gln His
65                  70                  75                  80
Ser Cys Thr Leu Asp Glu Cys Phe Gln Phe Tyr Thr Lys Glu Glu Gln
                85                  90                  95
Leu Ala Gln Asp Asp Ala Trp Lys Cys Pro His Cys Gln Val Leu Gln
            100                 105                 110
Gln Gly Met Val Lys Leu Ser Leu Trp Thr Leu Pro Asp Ile Leu Ile
        115                 120                 125
Ile His Leu Lys Arg Phe Cys Gln Val Gly Glu Arg Arg Asn Lys Leu
    130                 135                 140
Ser Thr Leu Val Lys Phe Pro Leu Ser Gly Leu Asn Met Ala Pro His
145                 150                 155                 160
Val Ala Gln Arg Ser Thr Ser Pro Glu Ala Gly Leu Gly Pro Trp Pro
                165                 170                 175
Ser Trp Lys Gln Pro Asp Cys Leu Pro Thr Ser Tyr Pro Leu Asp Phe
            180                 185                 190
Leu Tyr Asp Leu Tyr Ala Val Cys Asn His His Gly Asn Leu Gln Gly
        195                 200                 205
Gly His Tyr Thr Ala Tyr Cys Arg Asn Ser Leu Asp Gly Gln Trp Tyr
    210                 215                 220
```

Ser Tyr Asp Asp Ser Thr Val Glu Pro Leu Arg Glu Asp Glu Val Asn
225 230 235 240
Thr Arg Gly Ala Tyr Ile Leu Phe Tyr Gln Lys Arg Asn Ser Ile Pro
245 250 255

<210> 62
<211> 307
<212> PRT
<213> Homo sapiens

<400> 62
Leu Pro Pro Ala Phe Phe Leu Gly Leu Val Pro Gly Leu Val Asn Leu
1 5 10 15
Gly Asn Thr Cys Phe Met Asn Ser Leu Leu Gln Gly Leu Ser Ala Cys
20 25 30
Pro Ala Phe Ile Ser Leu Ala Leu Phe Ile Phe Glu Ser Leu Leu Pro
35 40 45
Leu Tyr Ser Cys Ser Phe Ile Ala Gln Glu Gly Ile His Leu Tyr Arg
50 55 60
Gln Gln Asp Ala His Glu Leu Phe His Val Ile Thr Ser Ser Leu Glu
65 70 75 80
Asp Glu Arg Asp Arg Gln Pro Arg Ser Gln His Pro Phe His Gly Arg
85 90 95
Leu Thr Ser Asn Met Val Cys Lys His Cys Glu His Gln Ser Pro Val
100 105 110
Arg Phe Asp Thr Phe Asp Ser Leu Ser Leu Ser Ile Pro Ala Ala Thr
115 120 125
Trp Gly His Pro Leu Thr Leu Asp His Cys Leu His His Phe Ile Ser
130 135 140
Ser Glu Ser Val Arg Asp Val Val Cys Asp Asn Cys Thr Lys Arg Thr
145 150 155 160
Thr Phe Val Lys Gln Leu Lys Leu Gly Lys Val Ser Pro His Tyr Thr
165 170 175
Pro Cys Trp Leu Cys Phe Glu Asp Ser Val Tyr Pro Ala Pro Glu Thr
180 185 190
Thr Arg Phe Ser Arg Phe Leu Phe His Pro Gln Leu Pro Gln Cys Leu
195 200 205
Cys Ile His Leu Gln Arg Leu Ser Trp Ser Ser His Gly Thr Pro Leu
210 215 220
Lys Arg His Glu His Val His Ser Ser Thr Tyr Leu Phe Arg Leu Met
225 230 235 240
Ala Val Val Val His His Gly Asp Met His Ser Gly His Phe Val Thr
245 250 255
Tyr Arg Arg Ser Pro Pro Ser Ala Arg Asn Pro Leu Ser Thr Ser Asn
260 265 270
Gln Trp Leu Trp Val Ser Asp Asp Thr Val Arg Lys Ala Ser Leu Gln
275 280 285
Glu Val Leu Ser Ser Ser Ala Tyr Leu Leu Phe Tyr Glu Arg Val Leu
290 295 300
Ser Arg Met
305

<210> 63
<211> 316
<212> PRT
<213> Homo sapiens

<400> 63
Met Leu Ser Ser Pro Asp Phe Tyr Pro Ala Tyr Pro Ser Ala Met Gln
1 5 10 15

```
Ser Leu Ser Leu Gly Ser Leu Ala Arg Ala Leu Glu Leu Met Thr Gln
            20                  25                  30
Tyr Phe Asn Asn Trp Asn Trp Val Tyr Asp Asn Ile Ile Asp Gln Asn
        35                  40                  45
Glu Ser Lys Leu Ser Lys Ser Arg Arg Glu Glu Ile Glu Arg Asp Arg
    50                  55                  60
Glu Arg Lys Glu Arg Arg Glu Gly Asp Arg Glu Lys Lys Arg Gln Asn
65                  70                  75                  80
Gly Val Val Glu Val Pro Phe Leu Leu Ser Ser Lys Tyr Asp Glu Pro
                85                  90                  95
Ser Arg Gln Val Ile Leu Glu Ala Leu Ala Glu Phe Glu Arg Ser Thr
            100                 105                 110
Cys Ile Arg Phe Val Thr Tyr Gln Asp Gln Arg Asp Phe Ile Ser Ile
        115                 120                 125
Ile Pro Met Tyr Gly Cys Phe Ser Ser Val Gly Arg Ser Gly Gly Met
    130                 135                 140
Gln Val Val Ser Leu Ala Pro Thr Cys Leu Gln Lys Gly Arg Gly Ile
145                 150                 155                 160
Val Leu His Glu Leu Met His Val Leu Gly Phe Trp His Glu His Thr
                165                 170                 175
Arg Ala Asp Arg Asp Arg Tyr Ile Arg Val Asn Trp Asn Glu Ile Leu
            180                 185                 190
Pro Gly Phe Glu Ile Asn Phe Ile Lys Ser Arg Ser Ser Asn Met Leu
        195                 200                 205
Thr Pro Tyr Asp Tyr Ser Ser Val Met His Tyr Gly Arg Leu Ala Phe
    210                 215                 220
Ser Arg Arg Gly Leu Pro Thr Ile Thr Pro Leu Trp Ala Pro Ser Val
225                 230                 235                 240
His Ile Gly Gln Arg Trp Asn Leu Ser Ala Ser Asp Ile Thr Arg Val
                245                 250                 255
Leu Gln Leu Tyr Gly Cys Ser Pro Lys Leu Glu Lys Val Asn Thr Val
            260                 265                 270
Asn Ile Lys Ile Ile Phe Leu Tyr Thr Gln Leu Glu Asn Thr Ile Val
        275                 280                 285
Ser Lys His Thr Ile Ile Ile Ala Ile Ile Lys Val Pro Arg Asn Lys
    290                 295                 300
Ser Asp Ser Cys Ile Lys Tyr Leu Arg Lys Asn Glu
305                 310                 315
```

<210> 64
<211> 725
<212> PRT
<213> Homo sapiens

```
<400> 64
Phe Gln Leu Trp Ile Trp Leu Arg Pro Cys Pro Val Thr Trp Ile Pro
 1               5                  10                  15
Arg Phe Pro Gly Gly Gly Val Phe Pro Gly Gly Ser Leu Ser Pro Leu
            20                  25                  30
His Ile Leu Gly Thr Lys Ala Phe Phe Lys Val Leu Phe Leu Asp His
        35                  40                  45
His Phe Arg Leu Tyr Met Glu His Ser Asn Asp Ile Ile Ser Pro His
    50                  55                  60
Phe Lys Glu Ile Thr Gln Ile Ile Thr Ser Phe Gln Glu Ile Ile Glu
65                  70                  75                  80
Glu Glu Phe Gly Ile Ser Gln Cys Tyr Thr Tyr Asn Asn Pro Ser Lys
                85                  90                  95
Ser Asp Ile Arg Ile His Trp Thr Val Ser Asp Leu Ser Gln Val Phe
            100                 105                 110
Ile Leu Ser Arg Ala Lys Ser Leu Tyr Ile Gln Ile Leu Ser Gln Arg
        115                 120                 125
```

```
His Ser Arg Lys Lys Arg Leu Ile Ser Tyr Pro Arg Tyr Ile Glu Ile
    130                 135                 140
Met Val Thr Ala Asp Ala Lys Val Val Ser Ala His Gly Ser Asn Leu
145                 150                 155                 160
Gln Asn Tyr Ile Leu Thr Leu Met Ser Ile Val Ala Thr Ile Tyr Lys
                165                 170                 175
Asp Pro Ser Ile Gly Asn Leu Ile His Ile Val Val Val Lys Leu Val
            180                 185                 190
Met Ile His Arg Glu Glu Glu Gly Pro Val Ile Asn Phe Asp Gly Ala
        195                 200                 205
Thr Thr Leu Lys Asn Phe Cys Ser Trp Gln Gln Thr Gln Asn Asp Leu
    210                 215                 220
Asp Asp Val His Pro Ser His His Asp Thr Ala Val Leu Ile Thr Arg
225                 230                 235                 240
Glu Asp Ile Cys Ser Ser Lys Glu Lys Cys Asn Met Leu Gly Leu Ser
                245                 250                 255
Tyr Leu Gly Thr Ile Cys Asp Pro Leu Gln Ser Cys Phe Ile Asn Glu
            260                 265                 270
Glu Lys Gly Leu Ile Ser Ala Phe Thr Ile Ala His Glu Leu Gly His
        275                 280                 285
Thr Leu Gly Val Gln His Asp Asp Asn Pro Arg Cys Lys Glu Met Lys
    290                 295                 300
Val Thr Lys Tyr His Val Met Ala Pro Ala Leu Ser Phe His Met Ser
305                 310                 315                 320
Pro Trp Ser Trp Ser Asn Cys Ser Arg Lys Tyr Val Thr Glu Phe Leu
                325                 330                 335
Asp Thr Gly Tyr Gly Glu Cys Leu Leu Asp Lys Pro Asp Glu Glu Ile
            340                 345                 350
Tyr Asn Leu Pro Ser Glu Leu Pro Gly Ser Arg Tyr Asp Gly Asn Lys
        355                 360                 365
Gln Cys Glu Leu Ala Phe Gly Pro Gly Ser Gln Met Cys Pro His Ile
    370                 375                 380
Glu Asn Ile Cys Met His Leu Trp Cys Thr Ser Thr Glu Lys Leu His
385                 390                 395                 400
Lys Gly Cys Phe Thr Gln His Val Pro Pro Ala Asp Gly Thr Asp Cys
                405                 410                 415
Gly Pro Gly Met His Cys Arg His Gly Leu Cys Val Asn Lys Glu Thr
            420                 425                 430
Glu Thr Arg Pro Val Asn Gly Glu Trp Gly Pro Trp Glu Pro Tyr Ser
        435                 440                 445
Ser Cys Ser Arg Thr Cys Gly Gly Gly Ile Glu Ser Ala Thr Arg Arg
    450                 455                 460
Cys Asn Arg Pro Glu Pro Arg Asn Gly Gly Asn Tyr Cys Val Gly Arg
465                 470                 475                 480
Arg Met Lys Phe Arg Ser Cys Asn Thr Asp Ser Cys Pro Lys Gly Thr
                485                 490                 495
Gln Asp Phe Arg Glu Lys Gln Cys Ser Asp Phe Asn Gly Lys His Leu
            500                 505                 510
Asp Ile Ser Gly Ile Pro Ser Asn Val Arg Trp Leu Pro Arg Tyr Ser
        515                 520                 525
Gly Ile Gly Thr Lys Asp Arg Cys Lys Leu Tyr Cys Gln Val Ala Gly
    530                 535                 540
Thr Asn Tyr Phe Tyr Leu Leu Lys Asp Met Val Glu Asp Gly Thr Pro
545                 550                 555                 560
Cys Gly Thr Glu Thr His Asp Ile Cys Val Gln Gly Gln Cys Met Ala
                565                 570                 575
Ala Gly Cys Asp His Val Leu Asn Ser Ser Ala Lys Ile Asp Lys Cys
            580                 585                 590
Gly Val Cys Gly Gly Asp Asn Ser Ser Cys Lys Thr Ile Thr Gly Val
        595                 600                 605
Phe Asn Ser Ser His Tyr Gly Tyr Asn Val Val Val Lys Ile Pro Ala
    610                 615                 620
```

```
Gly Ala Thr Asn Val Asp Ile Arg Gln Tyr Ser Tyr Ser Gly Gln Pro
625                 630                 635                 640
Asp Asp Ser Tyr Leu Ala Leu Ser Asp Ala Glu Gly Asn Phe Leu Phe
                645                 650                 655
Asn Gly Asn Phe Leu Leu Ser Thr Ser Lys Lys Glu Ile Asn Val Gln
            660                 665                 670
Gly Thr Arg Thr Val Ile Glu Tyr Ser Gly Ser Asn Asn Ala Val Glu
        675                 680                 685
Arg Ile Asn Ser Thr Asn Arg Gln Glu Lys Glu Leu Ile Leu Gln Val
    690                 695                 700
Leu Cys Val Gly Asn Leu Tyr Asn Pro Asp Val His Tyr Ser Phe Asn
705                 710                 715                 720
Ile Pro Leu Glu Glu
                725
```

<210> 65
<211> 700
<212> PRT
<213> Homo sapiens

<400> 65

```
Pro Ser Leu Leu Ser Cys Leu Leu Ser Phe Pro Arg Pro Gly Pro Asp
 1               5                  10                  15
Ile Ala Trp Gln Leu Ser Cys Lys Gly Ser Trp Ile Gly Thr Gln Thr
            20                  25                  30
His Ser Leu Thr Val Ser Ser Ala Thr Asp Phe Val Leu Arg Trp Gln
        35                  40                  45
Asn Arg Met Val Glu Tyr Pro Gly Val Pro Gln Met Pro Tyr Gly Gly
    50                  55                  60
His Ser Ser Pro Met Thr Phe Leu Leu Tyr Gly Asp Ile Ala Asn Phe
65                  70                  75                  80
Asp Phe Tyr Ser Asn Leu Val Val Thr Ala Pro Pro Val Gly Trp Thr
                85                  90                  95
Ser Leu Ser Ser Cys Leu Asp Leu Pro Asn Leu Leu Gly Leu Val Gly
            100                 105                 110
Asp Gln Leu Gly Asp Thr Glu Arg Lys Arg Arg His Ala Lys Pro Gly
        115                 120                 125
Ser Tyr Ser Ile Glu Val Leu Leu Val Val Asp Asp Ser Val Val Arg
    130                 135                 140
Phe His Gly Lys Glu His Val Gln Asn Tyr Val Leu Thr Leu Met Asn
145                 150                 155                 160
Ile Val Asn Glu Ile Tyr His Asp Glu Ser Leu Gly Val His Ile Asn
                165                 170                 175
Ile Ala Leu Val Arg Leu Ile Met Val Gly Tyr Arg Gln Ser Leu Ser
            180                 185                 190
Leu Ile Glu Arg Gly Asn Pro Ser Arg Ser Leu Glu Gln Val Cys Arg
        195                 200                 205
Trp Ala His Ser Gln Gln Arg Gln Asp Pro Ser His Ala Glu His His
    210                 215                 220
Asp His Val Val Phe Leu Thr Arg Gln Asp Phe Gly Pro Ser Gly Met
225                 230                 235                 240
Gln Gly Tyr Ala Pro Val Thr Gly Met Cys His Pro Leu Arg Ser Cys
                245                 250                 255
Ala Leu Asn His Glu Asp Gly Phe Ser Ser Ala Phe Val Ile Ala His
            260                 265                 270
Glu Thr Gly His Val Leu Gly Met Glu His Asp Gly Gln Gly Asn Gly
        275                 280                 285
Cys Ala Asp Glu Thr Ser Leu Gly Ser Val Met Ala Pro Leu Val Gln
    290                 295                 300
Ala Ala Phe His Arg Phe His Trp Ser Arg Cys Ser Lys Leu Glu Leu
305                 310                 315                 320
```

```
Ser Arg Tyr Leu Pro Ser Tyr Asp Cys Leu Leu Asp Asp Pro Phe Asp
                325                 330                 335
Pro Ala Trp Pro Gln Pro Pro Glu Leu Pro Gly Ile Asn Tyr Ser Met
            340                 345                 350
Asp Glu Gln Cys Arg Phe Asp Phe Gly Ser Gly Tyr Gln Thr Cys Leu
        355                 360                 365
Ala Phe Arg Thr Phe Glu Pro Cys Lys Gln Leu Trp Cys Ser His Pro
    370                 375                 380
Asp Asn Pro Tyr Phe Cys Lys Thr Lys Lys Gly Pro Pro Leu Asp Gly
385                 390                 395                 400
Thr Glu Cys Ala Pro Gly Lys Trp Cys Phe Lys Gly His Cys Ile Trp
                405                 410                 415
Lys Ser Pro Glu Gln Thr Tyr Gly Gln Asp Gly Gly Trp Ser Ser Trp
            420                 425                 430
Thr Lys Phe Gly Ser Cys Ser Arg Ser Cys Gly Gly Gly Val Arg Ser
        435                 440                 445
Arg Ser Arg Ser Cys Asn Asn Pro Ser Pro Ala Tyr Gly Gly Arg Leu
    450                 455                 460
Cys Leu Gly Pro Met Phe Glu Tyr Gln Val Cys Asn Ser Glu Glu Cys
465                 470                 475                 480
Pro Gly Thr Tyr Glu Asp Phe Arg Ala Gln Gln Cys Ala Lys Arg Asn
                485                 490                 495
Ser Tyr Tyr Val His Gln Asn Ala Lys His Ser Trp Val Pro Tyr Glu
            500                 505                 510
Pro Asp Asp Asp Ala Gln Lys Cys Glu Leu Ile Cys Gln Ser Ala Asp
        515                 520                 525
Thr Gly Asp Val Val Phe Met Asn Gln Val Val His Asp Gly Thr Arg
    530                 535                 540
Cys Ser Tyr Arg Asp Pro Tyr Ser Val Cys Ala Arg Gly Glu Cys Val
545                 550                 555                 560
Pro Val Gly Cys Asp Lys Glu Val Gly Ser Met Lys Ala Asp Asp Lys
                565                 570                 575
Cys Gly Val Cys Gly Gly Asp Asn Ser His Cys Arg Thr Val Lys Gly
            580                 585                 590
Thr Leu Gly Lys Ala Ser Lys Gln Ala Ala Leu Lys Leu Val Gln Ile
        595                 600                 605
Pro Ala Gly Ala Arg His Ile Gln Ile Glu Ala Leu Glu Lys Ser Pro
    610                 615                 620
His Arg Ile Val Val Lys Asn Gln Val Thr Gly Ser Phe Ile Leu Asn
625                 630                 635                 640
Pro Lys Gly Lys Glu Ala Thr Ser Arg Thr Phe Thr Ala Met Gly Leu
                645                 650                 655
Glu Trp Glu Asp Ala Val Glu Asp Ala Lys Glu Ser Leu Lys Thr Ser
            660                 665                 670
Gly Pro Leu Pro Glu Ala Ile Ala Ile Leu Val Ser Pro Thr Leu Asp
        675                 680                 685
Thr Gln Asn Ile Lys Glu Pro Arg His Arg Pro Asp
    690                 695                 700
```

<210> 66
<211> 749
<212> PRT
<213> Homo sapiens

<400> 66

```
Leu Leu Leu Leu Leu Leu Leu Leu Leu Phe Leu Arg Gln Ser Trp Gln
 1               5                  10                  15
Gly Pro Ile Ile Ser Ala Thr Gln Glu Ala Glu Ala Val Glu Ser Leu
            20                  25                  30
Glu Pro Arg Arg Arg Arg Leu His Ser Gly Val Arg Asp Gln Pro Gly
        35                  40                  45
```

```
Gln His Gly Glu Thr Leu Ser Leu Leu Lys Ile Gln Lys Leu Asp Arg
    50                  55                  60
His Gly Leu Ile Arg Thr Arg Lys Asn Glu Phe Leu Ile Ser Pro Leu
65                  70                  75                  80
Pro Gln Leu Leu Ala Gln Glu His Asn Tyr Ser Ser Pro Ala Gly His
                85                  90                  95
His Pro His Val Leu Leu Asp Cys Phe Tyr His Cys His Ile Lys Asp
            100                 105                 110
Phe Ser Ser Ser Leu Val Ser Val Ser Leu Ser Thr Val Leu Ser Arg
        115                 120                 125
Tyr Ser Leu Ile Ser Leu Pro Asn Ala Leu Leu Phe Ile Val Asp Ala
    130                 135                 140
Pro Lys Pro Pro Thr Glu Asp Thr Tyr Leu Arg Phe Asp Glu Tyr Gly
145                 150                 155                 160
Ser Ser Gly Arg Pro Arg Arg Ser Ala Gly Lys Ser Gln Lys Gly Leu
                165                 170                 175
Asn Val Glu Thr Leu Val Val Ala Asp Lys Lys Met Val Glu Lys His
            180                 185                 190
Gly Lys Gly Asn Val Thr Thr Tyr Ile Leu Thr Val Met Asn Met Val
        195                 200                 205
Ser Gly Leu Phe Lys Asp Gly Thr Ile Gly Ser Asp Ile Asn Val Val
    210                 215                 220
Val Val Ser Leu Ile Leu Leu Glu Gln Glu Pro Gly Gly Leu Leu Ile
225                 230                 235                 240
Asn His His Ala Asp Gln Ser Leu Asn Ser Phe Cys Gln Trp Gln Ser
                245                 250                 255
Ala Leu Ile Gly Lys Asn Gly Lys Arg His Asp His Ala Ile Leu Leu
            260                 265                 270
Thr Gly Phe Asp Ile Cys Ser Trp Lys Asn Glu Pro Cys Asp Thr Leu
        275                 280                 285
Gly Phe Ala Pro Ile Ser Gly Met Cys Ser Lys Tyr Arg Ser Cys Thr
    290                 295                 300
Ile Asn Glu Asp Thr Gly Leu Gly Leu Ala Phe Thr Ile Ala His Glu
305                 310                 315                 320
Ser Gly His Asn Phe Gly Met Ile His Asp Gly Glu Gly Asn Pro Cys
                325                 330                 335
Arg Lys Ala Glu Gly Asn Ile Met Ser Pro Thr Leu Thr Gly Asn Asn
            340                 345                 350
Gly Val Phe Ser Trp Ser Ser Cys Ser Arg Gln Tyr Leu Lys Lys Phe
        355                 360                 365
Leu Ser Thr Pro Gln Ala Gly Cys Leu Val Asp Glu Pro Lys Gln Ala
    370                 375                 380
Gly Gln Tyr Lys Tyr Pro Asp Lys Leu Pro Gly Gln Ile Tyr Asp Ala
385                 390                 395                 400
Asp Thr Gln Cys Lys Trp Gln Phe Gly Ala Lys Ala Lys Leu Cys Ser
                405                 410                 415
Leu Gly Phe Asp Ile Cys Lys Ser Leu Trp Cys His Arg Val Gly His
            420                 425                 430
Arg Cys Glu Thr Lys Phe Met Pro Ala Ala Glu Gly Thr Val Cys Gly
        435                 440                 445
Leu Ser Met Trp Cys Arg Gln Gly Gln Cys Val Lys Phe Gly Glu Leu
    450                 455                 460
Gly Pro Arg Pro Ile His Gly Gln Trp Ser Ala Trp Ser Lys Trp Ser
465                 470                 475                 480
Glu Cys Ser Arg Thr Cys Gly Gly Gly Val Lys Phe Gln Glu Arg His
                485                 490                 495
Cys Asn Asn Pro Lys Pro Gln Tyr Gly Gly Leu Phe Cys Pro Gly Ser
            500                 505                 510
Ser Arg Ile Tyr Gln Leu Cys Asn Ile Asn Pro Cys Asn Glu Asn Ser
        515                 520                 525
Leu Asp Phe Arg Ala Gln Gln Cys Ala Glu Tyr Asn Ser Lys Pro Phe
    530                 535                 540
```

```
Arg Gly Trp Phe Tyr Gln Trp Lys Pro Tyr Thr Lys Val Glu Glu Glu
545                 550                 555                 560
Asp Arg Cys Lys Leu Tyr Cys Lys Ala Glu Asn Phe Glu Phe Phe Phe
                565                 570                 575
Ala Met Ser Gly Lys Val Lys Asp Gly Thr Pro Cys Ser Pro Asn Lys
            580                 585                 590
Asn Asp Val Cys Ile Asp Gly Val Cys Glu Leu Val Gly Cys Asp His
        595                 600                 605
Glu Leu Gly Ser Lys Ala Val Ser Asp Ala Cys Gly Val Cys Lys Gly
    610                 615                 620
Asp Asn Ser Thr Cys Lys Phe Tyr Lys Gly Leu Phe Lys Gln Phe Ser
625                 630                 635                 640
Cys Leu Thr Leu Leu Lys Tyr Tyr Pro Val Val Leu Ile Pro Ala Gly
                645                 650                 655
Ala Arg Ser Ile Glu Ile Gln Glu Leu Gln Val Ser Ser Ser Tyr Leu
            660                 665                 670
Ala Val Arg Ser Leu Ser Gln Lys Tyr Tyr Leu Thr Gly Gly Trp Ser
        675                 680                 685
Ile Asp Trp Pro Gly Glu Phe Pro Phe Ala Gly Thr Thr Phe Glu Tyr
    690                 695                 700
Gln Arg Ser Phe Asn Arg Pro Glu Arg Leu Tyr Ala Pro Gly Pro Thr
705                 710                 715                 720
Asn Glu Thr Leu Cys Cys Ser Val Ala Gln Ala Gly Gly Gln Leu Arg
                725                 730                 735
Asp Leu Gly Ser Leu Gln Ala Pro Pro Pro Glu Phe Thr
            740                 745
```

<210> 67
<211> 722
<212> PRT
<213> Homo sapiens

<400> 67

```
Leu Leu Leu Trp Arg Cys Pro Leu Ser Pro Ala Phe Pro Leu Leu Pro
 1               5                  10                  15
Ser Arg Leu Arg Leu Ser Ala Pro Leu Thr Ser Pro Pro Pro Leu Cys
            20                  25                  30
Ser Leu Ser Leu His Ser Pro Leu Leu Gly Pro Leu Thr Pro Pro Pro
        35                  40                  45
Ser Pro Pro Pro Leu Leu Ser Pro Leu Pro Ala Pro Arg Ser Pro Thr
    50                  55                  60
Ala Pro Ala Ala Pro Ala Ala Ala Ala Ala Thr Pro Pro Pro Ala Pro
65                  70                  75                  80
His Gly Ala Ser Pro Leu Leu Thr Leu Leu Ile Ser Glu Tyr Asp Leu
                85                  90                  95
Val Ser Ala Tyr Glu Val Asp His Arg Gly Asp Tyr Val Ser His Glu
            100                 105                 110
Ile Met His His Gln Arg Arg Arg Arg Ala Val Ala Val Ser Glu Val
        115                 120                 125
Glu Ser Leu His Leu Arg Leu Lys Gly Pro Arg His Asp Phe His Met
    130                 135                 140
Asp Leu Arg Thr Ser Ser Ser Leu Val Ala Pro Gly Phe Ile Val Gln
145                 150                 155                 160
Thr Leu Gly Lys Thr Gly Thr Lys Ser Val Gln Thr Leu Pro Pro Glu
                165                 170                 175
Asp Phe Cys Phe Tyr Gln Gly Ser Leu Arg Ser His Arg Asn Ser Ser
            180                 185                 190
Val Ala Leu Ser Thr Cys Gln Gly Leu Ser Gly Met Ile Arg Thr Glu
        195                 200                 205
Glu Ala Asp Tyr Phe Leu Arg Pro Leu Pro Ser His Leu Ser Trp Lys
    210                 215                 220
```

```
Leu Gly Arg Ala Ala Gln Gly Ser Ser Pro Ser His Val Leu Asn Glu
225                 230                 235                 240
Glu Leu Asn Val Glu Thr Leu Val Val Val Asp Lys Lys Met Met Gln
                245                 250                 255
Asn His Gly His Glu Asn Ile Thr Thr Tyr Val Leu Thr Ile Leu Asn
            260                 265                 270
Met Val Ser Ala Leu Phe Lys Asp Gly Thr Ile Gly Gly Asn Ile Asn
        275                 280                 285
Ile Ala Ile Val Gly Leu Ile Leu Leu Glu Asp Glu Gln Pro Gly Leu
    290                 295                 300
Val Ile Ser His His Ala Asp His Thr Leu Ser Ser Phe Cys Gln Trp
305                 310                 315                 320
Gln Ser Gly Leu Met Gly Lys Asp Gly Thr Arg His Asp His Ala Ile
                325                 330                 335
Leu Leu Thr Gly Leu Asp Ile Cys Ser Trp Lys Asn Glu Pro Cys Asp
            340                 345                 350
Thr Leu Gly Phe Ala Pro Ile Ser Gly Met Cys Ser Lys Tyr Arg Ser
        355                 360                 365
Cys Thr Ile Asn Glu Asp Thr Gly Leu Gly Leu Ala Phe Thr Ile Ala
    370                 375                 380
His Glu Ser Gly His Asn Phe Gly Met Ile His Asp Gly Glu Gly Asn
385                 390                 395                 400
Met Cys Lys Lys Ser Glu Gly Asn Ile Met Ser Pro Thr Leu Ala Gly
                405                 410                 415
Arg Asn Gly Val Phe Ser Trp Ser Pro Cys Ser Arg Gln Tyr Leu His
            420                 425                 430
Lys Phe Leu Ser Thr Ala Gln Ala Ile Cys Leu Ala Asp Gln Pro Lys
        435                 440                 445
Pro Val Lys Glu Tyr Lys Tyr Pro Glu Lys Leu Pro Gly Glu Leu Tyr
    450                 455                 460
Asp Ala Asn Thr Gln Cys Lys Trp Gln Phe Gly Glu Lys Ala Lys Leu
465                 470                 475                 480
Cys Met Leu Asp Phe Lys Lys Ala Thr Leu Trp Cys His Arg Ile Gly
                485                 490                 495
Arg Lys Cys Glu Thr Lys Phe Met Pro Ala Ala Glu Gly Thr Ile Cys
            500                 505                 510
Gly His Asp Met Trp Cys Arg Gly Gly Gln Cys Val Lys Tyr Gly Asp
        515                 520                 525
Glu Gly Pro Lys Pro Thr His Gly His Trp Ser Asp Trp Ser Ser Trp
    530                 535                 540
Ser Pro Cys Ser Arg Thr Cys Gly Gly Gly Val Ser His Arg Ser Arg
545                 550                 555                 560
Leu Cys Thr Asn Pro Lys Pro Ser His Gly Gly Lys Phe Cys Glu Gly
                565                 570                 575
Ser Thr Arg Thr Leu Lys Leu Cys Asn Ser Gln Lys Cys Pro Arg Asp
            580                 585                 590
Ser Val Asp Phe Arg Ala Ala Gln Cys Ala Glu His Asn Ser Arg Arg
        595                 600                 605
Phe Arg Gly Arg His Tyr Lys Trp Lys Pro Tyr Thr Gln Val Glu Asp
    610                 615                 620
Gln Asp Leu Cys Lys Leu Tyr Cys Ile Ala Glu Gly Phe Asp Phe Phe
625                 630                 635                 640
Phe Ser Leu Ser Asn Lys Val Lys Asp Gly Thr Pro Cys Ser Glu Asp
                645                 650                 655
Ser Arg Asn Val Cys Ile Asp Gly Ile Cys Glu Arg Val Gly Cys Asp
            660                 665                 670
Asn Val Leu Gly Ser Asp Ala Val Glu Asp Val Cys Gly Val Cys Asn
        675                 680                 685
Gly Asn Asn Ser Ala Cys Thr Ile His Arg Gly Leu Tyr Thr Lys His
    690                 695                 700
His His Thr Asn Arg Glu Tyr Phe Arg Ala Ala Cys Lys Pro Trp Ala
705                 710                 715                 720
```

Lys Lys

<210> 68
<211> 743
<212> PRT
<213> Homo sapiens

<400> 68

```
Met Ser Ser Tyr Arg Pro Arg Val Ile Ser Arg Met Pro His Leu Leu
 1               5                  10                  15
Arg Leu Leu Leu Ala Val Thr Val Ser Gln Thr Phe Ile Val Phe Leu
            20                  25                  30
Phe Val Phe Leu Phe Phe Val Ile Leu Thr Val Leu Arg Ile Thr Asp
        35                  40                  45
Gln Val Ser Glu Val Cys Thr Thr Pro Gly Cys Val Ile Ala Ala Ala
    50                  55                  60
Ala Arg Ile Leu Gln Asn Met Asp Pro Thr Thr Glu Pro Cys Asp Asp
65                  70                  75                  80
Phe Tyr Gln Phe Ala Cys Gly Gly Trp Leu Arg Arg His Val Ile Pro
                85                  90                  95
Glu Thr Asn Ser Arg Tyr Ser Ile Phe Asp Val Leu Arg Asp Glu Leu
            100                 105                 110
Glu Val Ile Leu Lys Ala Val Leu Glu Asn Ser Thr Ala Lys Asp Arg
        115                 120                 125
Pro Ala Val Glu Lys Ala Arg Thr Leu Tyr Arg Ser Cys Met Asn Gln
    130                 135                 140
Ser Val Ile Glu Lys Arg Gly Ser Gln Pro Leu Leu Asp Ile Leu Glu
145                 150                 155                 160
Val Val Gly Gly Trp Pro Val Ala Met Asp Arg Trp Asn Glu Thr Val
                165                 170                 175
Gly Leu Glu Trp Glu Leu Glu Arg Gln Leu Ala Leu Met Asn Ser Gln
            180                 185                 190
Phe Asn Arg Arg Val Leu Ile Asp Leu Phe Ile Trp Asn Asp Asp Gln
        195                 200                 205
Asn Ser Ser Arg His Ile Ile Tyr Ile Asp Gln Pro Thr Leu Gly Met
    210                 215                 220
Pro Ser Arg Glu Tyr Tyr Phe Asn Gly Gly Ser Asn Arg Lys Val Arg
225                 230                 235                 240
Glu Ala Tyr Leu Gln Phe Met Val Ser Val Ala Thr Leu Leu Arg Glu
                245                 250                 255
Asp Ala Asn Leu Pro Arg Asp Ser Cys Leu Val Gln Glu Asp Met Val
            260                 265                 270
Gln Val Leu Glu Leu Glu Thr Gln Leu Ala Lys Ala Thr Val Pro Gln
        275                 280                 285
Glu Glu Arg His Asp Val Ile Ala Leu Tyr His Arg Met Gly Leu Glu
    290                 295                 300
Glu Leu Gln Ser Gln Phe Gly Leu Lys Gly Phe Asn Trp Thr Leu Phe
305                 310                 315                 320
Ile Gln Thr Val Leu Ser Ser Val Lys Ile Lys Leu Leu Pro Asp Glu
                325                 330                 335
Glu Val Val Val Tyr Gly Ile Pro Tyr Leu Gln Asn Leu Glu Asn Ile
            340                 345                 350
Ile Asp Thr Tyr Ser Ala Arg Thr Ile Gln Asn Tyr Leu Val Trp Arg
        355                 360                 365
Leu Val Leu Asp Arg Ile Gly Ser Leu Ser Gln Arg Phe Lys Asp Thr
    370                 375                 380
Arg Val Asn Tyr Arg Lys Ala Leu Phe Gly Thr Met Val Glu Glu Val
385                 390                 395                 400
Arg Trp Arg Glu Cys Val Gly Tyr Val Asn Ser Asn Met Glu Asn Ala
                405                 410                 415
```

```
Val Gly Ser Leu Tyr Val Arg Glu Ala Phe Pro Gly Asp Ser Lys Ser
            420                 425                 430
Met Val Arg Glu Leu Ile Asp Lys Val Arg Thr Val Phe Val Glu Thr
        435                 440                 445
Leu Asp Glu Leu Gly Trp Met Asp Glu Glu Ser Lys Lys Lys Ala Gln
    450                 455                 460
Glu Lys Ala Met Ser Ile Arg Glu Gln Ile Gly His Pro Asp Tyr Ile
465                 470                 475                 480
Leu Glu Glu Met Asn Arg Arg Leu Asp Glu Glu Tyr Ser Asn Leu Asn
                485                 490                 495
Phe Ser Glu Asp Leu Tyr Phe Glu Asn Ser Leu Gln Asn Leu Lys Val
            500                 505                 510
Gly Ala Gln Arg Ser Leu Arg Lys Leu Arg Glu Lys Val Asp Pro Asn
        515                 520                 525
Leu Trp Ile Ile Gly Ala Ala Val Val Asn Ala Phe Tyr Ser Pro Asn
    530                 535                 540
Arg Asn Gln Ile Val Phe Pro Ala Gly Ile Leu Gln Pro Pro Phe Phe
545                 550                 555                 560
Ser Lys Glu Gln Pro Gln Ala Leu Asn Phe Gly Gly Ile Gly Met Val
                565                 570                 575
Ile Gly His Glu Ile Thr His Gly Phe Asp Asp Asn Gly Arg Asn Phe
            580                 585                 590
Asp Lys Asn Gly Asn Met Met Asp Trp Trp Ser Asn Phe Ser Thr Gln
        595                 600                 605
His Phe Arg Glu Gln Ser Glu Cys Met Ile Tyr Gln Tyr Gly Asn Tyr
    610                 615                 620
Ser Trp Asp Leu Ala Asp Glu Gln Asn Val Asn Gly Phe Asn Thr Leu
625                 630                 635                 640
Gly Glu Asn Ile Ala Asp Asn Gly Gly Val Arg Gln Ala Tyr Lys Ala
                645                 650                 655
Tyr Leu Lys Trp Met Ala Glu Gly Gly Lys Asp Gln Gln Leu Pro Gly
            660                 665                 670
Leu Asp Leu Thr His Glu Gln Leu Phe Phe Ile Asn Tyr Ala Gln Val
        675                 680                 685
Trp Cys Gly Ser Tyr Arg Pro Glu Phe Ala Ile Gln Ser Ile Lys Thr
    690                 695                 700
Asp Val His Ser Pro Leu Lys Val Leu Gly Ser Leu Gln Asn Leu Ala
705                 710                 715                 720
Ala Phe Ala Asp Thr Phe His Cys Ala Arg Gly Thr Pro Met His Pro
                725                 730                 735
Lys Glu Arg Cys Arg Val Trp
            740
```

<210> 69
<211> 371
<212> PRT
<213> Homo sapiens

```
<400> 69
Met Thr Ala Leu Asp Arg Ala Cys Leu Tyr Trp Leu Phe Leu Phe Lys
 1               5                  10                  15
Leu Leu Val Ile Asp Ile Lys Asn Asn Gly His Phe Tyr Val Thr Leu
            20                  25                  30
Ala Asn Ser Lys His Leu Ser Leu Asp Phe Ile Val His Ile Thr Ile
        35                  40                  45
Ser Ile Leu Val Lys Ala Ile Gln Arg Val Ser Arg Lys Phe Gln Thr
    50                  55                  60
Phe Pro His Phe Pro Val Phe Tyr Trp Ala Leu Gln Thr Ile Tyr Glu
65                  70                  75                  80
Trp Met Arg Glu Ile Ser Glu Lys Tyr Lys Glu Val Val Thr Gln His
                85                  90                  95
```

```
Phe Leu Gly Val Thr Tyr Glu Thr His Pro Met Tyr Tyr Leu Lys Ile
            100                 105                 110
Ser Gln Pro Ser Gly Asn Pro Lys Lys Ile Ile Trp Met Asp Cys Gly
        115                 120                 125
Ile His Ala Arg Glu Trp Ile Ala Pro Ala Phe Cys Gln Trp Phe Val
    130                 135                 140
Lys Glu Ile Leu Gln Asn His Lys Asp Asn Ser Ser Ile Arg Lys Leu
145                 150                 155                 160
Leu Arg Asn Leu Asp Phe Tyr Val Leu Pro Val Leu Asn Ile Asp Gly
                165                 170                 175
Tyr Ile Tyr Thr Trp Thr Thr Asp Arg Leu Trp Arg Lys Ser Arg Ser
            180                 185                 190
Pro His Asn Asn Gly Thr Cys Phe Gly Thr Asp Leu Asn Arg Asn Phe
        195                 200                 205
Asn Ala Ser Trp Cys Ser Ile Gly Ala Ser Arg Asn Cys Gln Asp Gln
    210                 215                 220
Thr Phe Cys Gly Thr Gly Pro Val Ser Glu Pro Glu Thr Lys Ala Val
225                 230                 235                 240
Ala Ser Phe Ile Glu Ser Lys Lys Asp Asp Ile Leu Cys Phe Leu Thr
                245                 250                 255
Met His Ser Tyr Gly Gln Leu Ile Leu Thr Pro Tyr Gly Tyr Thr Lys
            260                 265                 270
Asn Lys Ser Ser Asn His Pro Glu Met Ile Gln Val Gly Gln Lys Ala
        275                 280                 285
Ala Asn Ala Leu Lys Ala Lys Tyr Gly Thr Asn Tyr Arg Val Gly Ser
    290                 295                 300
Ser Ala Asp Ile Leu Tyr Ala Ser Ser Gly Ser Ser Arg Asp Trp Ala
305                 310                 315                 320
Arg Asp Ile Gly Ile Pro Phe Ser Tyr Thr Phe Glu Leu Arg Asp Ser
                325                 330                 335
Gly Thr Tyr Gly Phe Val Leu Pro Glu Ala Gln Ile Gln Pro Thr Cys
            340                 345                 350
Glu Glu Thr Met Glu Ala Val Leu Ser Val Leu Asp Asp Leu Gln Lys
        355                 360                 365
Asn Pro Tyr
    370
```

<210> 70
<211> 318
<212> PRT
<213> Homo sapiens

```
<400> 70
Pro Asn Thr Gln Asn His Met Pro Leu Cys Leu Glu Leu Gly Ile Arg
 1               5                  10                  15
Ser Tyr His Ser Gly Phe Cys Gln Asp Cys Phe Arg Arg Asn Glu Asp
            20                  25                  30
Ile Ser His Ser Ile Val Leu Pro Ala Ala Val Ser Ser Ala His Pro
        35                  40                  45
Val Pro Lys His Ile Lys Lys Pro Asp Tyr Val Thr Thr Gly Ile Val
    50                  55                  60
Pro Asp Trp Gly Asp Ser Ile Glu Val Lys Asn Glu Asp Gln Ile Gln
65                  70                  75                  80
Gly Leu His Gln Ala Cys Gln Leu Ala Arg His Val Leu Leu Leu Ala
                85                  90                  95
Gly Lys Ser Leu Lys Val Asp Met Thr Thr Glu Glu Ile Asp Ala Leu
            100                 105                 110
Val His Arg Glu Ile Ile Ser His Asn Ala Tyr Pro Ser Pro Leu Gly
        115                 120                 125
Tyr Gly Gly Phe Pro Lys Ser Val Cys Thr Ser Val Asn Asn Val Leu
    130                 135                 140
```

```
Cys His Gly Ile Pro Asp Ser Arg Pro Leu Gln Asp Gly Asp Ile Ile
145                 150                 155                 160
Asn Ile Asp Val Thr Val Tyr Tyr Asn Gly Tyr His Gly Asp Thr Ser
                165                 170                 175
Glu Thr Phe Leu Val Gly Asn Val Asp Glu Cys Gly Lys Lys Leu Val
            180                 185                 190
Glu Val Ala Arg Arg Cys Arg Asp Glu Ala Ile Ala Ala Cys Arg Ala
        195                 200                 205
Gly Ala Pro Phe Ser Val Ile Gly Asn Thr Ile Ser His Ile Thr His
    210                 215                 220
Gln Asn Gly Phe Gln Val Cys Pro His Phe Val Gly His Gly Ile Gly
225                 230                 235                 240
Ser Tyr Phe His Gly His Pro Glu Ile Trp His His Ala Asn Asp Ser
                245                 250                 255
Asp Leu Pro Met Glu Glu Gly Met Ala Phe Thr Ile Glu Pro Ile Ile
            260                 265                 270
Thr Glu Gly Ser Pro Glu Phe Lys Val Leu Glu Asp Ala Trp Thr Val
        275                 280                 285
Val Ser Leu Asp Asn Gln Arg Ser Ala Gln Phe Glu His Thr Val Leu
    290                 295                 300
Ile Thr Ser Arg Gly Ala Gln Ile Leu Thr Lys Leu Pro His
305                 310                 315
```

<210> 71
<211> 237
<212> PRT
<213> Homo sapiens

<400> 71

```
Arg Ile Val Gly Gly Met Glu Ala Ser Pro Gly Glu Phe Pro Trp Gln
 1                5                  10                  15
Ala Ser Leu Arg Glu Asn Lys Glu His Phe Cys Gly Ala Ala Ile Ile
            20                  25                  30
Asn Ala Arg Trp Leu Val Ser Ala Ala His Cys Phe Asn Glu Phe Gln
        35                  40                  45
Asp Pro Thr Lys Trp Val Ala Tyr Val Gly Ala Thr Tyr Leu Ser Gly
    50                  55                  60
Ser Glu Ala Ser Thr Val Arg Ala Gln Val Val Gln Ile Val Lys His
65                  70                  75                  80
Pro Leu Tyr Asn Ala Asp Thr Ala Asp Phe Asp Val Ala Val Leu Glu
                85                  90                  95
Leu Thr Ser Pro Leu Pro Phe Gly Arg His Ile Gln Pro Val Cys Leu
            100                 105                 110
Pro Ala Ala Thr His Ile Phe Pro Pro Ser Lys Lys Cys Leu Ile Ser
        115                 120                 125
Gly Trp Gly Tyr Leu Lys Glu Asp Phe Leu Val Lys Pro Glu Val Leu
    130                 135                 140
Gln Lys Ala Thr Val Glu Leu Leu Asp Gln Ala Leu Cys Ala Ser Leu
145                 150                 155                 160
Tyr Gly His Ser Leu Thr Asp Arg Met Val Cys Ala Gly Tyr Leu Asp
                165                 170                 175
Gly Lys Val Asp Ser Cys Gln Gly Asp Ser Gly Gly Pro Leu Val Cys
            180                 185                 190
Glu Glu Pro Ser Gly Arg Phe Phe Leu Ala Gly Ile Val Ser Trp Gly
        195                 200                 205
Ile Gly Cys Ala Glu Ala Arg Arg Pro Gly Val Tyr Ala Arg Val Thr
    210                 215                 220
Arg Leu Arg Asp Trp Ile Leu Glu Ala Thr Thr Lys Ala
225                 230                 235
```

<210> 72
<211> 238
<212> PRT
<213> Homo sapiens

<400> 72
Arg Ile Val Gly Gly Ala Val Ser Ser Glu Gly Glu Trp Pro Trp Gln
1 5 10 15
Ala Ser Leu Gln Val Arg Gly Arg His Ile Cys Gly Gly Ala Leu Ile
20 25 30
Ala Asp Arg Trp Val Ile Thr Ala Ala His Cys Phe Gln Glu Asp Ser
35 40 45
Met Ala Ser Thr Val Leu Trp Thr Val Phe Leu Gly Lys Val Trp Gln
50 55 60
Asn Ser Arg Trp Pro Gly Glu Val Ser Phe Lys Val Ser Arg Leu Leu
65 70 75 80
Leu His Pro Tyr His Glu Glu Asp Ser His Asp Tyr Asp Val Ala Leu
85 90 95
Leu Gln Leu Asp His Pro Val Val Arg Ser Ala Ala Val Arg Pro Val
100 105 110
Cys Leu Pro Ala Arg Ser His Phe Phe Glu Pro Gly Leu His Cys Trp
115 120 125
Ile Thr Gly Trp Gly Ala Leu Arg Glu Gly Gly Pro Ile Ser Asn Ala
130 135 140
Leu Gln Lys Val Asp Val Gln Leu Ile Pro Gln Asp Leu Cys Ser Glu
145 150 155 160
Val Tyr Arg Tyr Gln Val Thr Pro Arg Met Leu Cys Ala Gly Tyr Arg
165 170 175
Lys Gly Lys Lys Asp Ala Cys Gln Gly Asp Ser Gly Gly Pro Leu Val
180 185 190
Cys Lys Ala Leu Ser Gly Arg Trp Phe Leu Ala Gly Leu Val Ser Trp
195 200 205
Gly Leu Gly Cys Gly Arg Pro Asn Tyr Phe Gly Val Tyr Thr Arg Ile
210 215 220
Thr Gly Val Ile Ser Trp Ile Gln Gln Thr Met Ala Gln Ser
225 230 235

<210> 73
<211> 235
<212> PRT
<213> Homo sapiens

<400> 73
Arg Ile Val Gly Gly Ser Ala Ala Gly Arg Gly Glu Trp Pro Trp Gln
1 5 10 15
Val Ser Leu Trp Leu Arg Arg Arg Glu His Arg Cys Gly Ala Val Leu
20 25 30
Val Ala Glu Arg Trp Leu Leu Ser Ala Ala His Cys Phe Asp Val Tyr
35 40 45
Gly Asp Pro Lys Gln Trp Ala Ala Phe Leu Gly Thr Pro Phe Leu Ser
50 55 60
Gly Ala Glu Gly Gln Leu Glu Arg Val Ala Arg Ile Tyr Lys His Pro
65 70 75 80
Phe Tyr Asn Leu Tyr Thr Leu Asp Tyr Asp Val Ala Leu Leu Glu Leu
85 90 95
Ala Gly Pro Val Arg Arg Ser Arg Leu Val Arg Pro Ile Cys Leu Pro
100 105 110
Glu Pro Ala Pro Arg Pro Pro Asp Gly Thr Arg Cys Val Ile Thr Gly
115 120 125
Trp Gly Ser Val Arg Glu Gly Gly Ser Met Ala Arg Gln Leu Gln Lys
130 135 140

```
Ala Ala Val Arg Leu Leu Ser Glu Gln Thr Cys Arg Arg Phe Tyr Pro
145                 150                 155                 160
Val Gln Ile Ser Ser Arg Met Leu Cys Ala Gly Phe Pro Gln Gly Gly
                165                 170                 175
Val Asp Ser Cys Ser Gly Asp Ala Gly Gly Pro Leu Ala Cys Arg Glu
            180                 185                 190
Pro Ser Gly Arg Trp Val Leu Thr Gly Val Thr Ser Trp Gly Tyr Gly
        195                 200                 205
Cys Gly Arg Pro His Phe Pro Gly Val Tyr Thr Arg Val Ala Ala Val
    210                 215                 220
Arg Gly Trp Ile Gly Gln His Ile Gln Asp Asn
225                 230                 235
```

<210> 74
<211> 238
<212> PRT
<213> Homo sapiens

<400> 74

```
Arg Ile Ile Gly Gly Thr Asp Thr Leu Glu Gly Gly Trp Pro Trp Gln
 1               5                  10                  15
Val Ser Leu His Phe Val Gly Ser Ala Tyr Cys Gly Ala Ser Val Ile
            20                  25                  30
Ser Arg Glu Trp Leu Leu Ser Ala Ala His Cys Phe His Gly Asn Arg
        35                  40                  45
Leu Ser Asp Pro Thr Pro Trp Thr Ala His Leu Gly Met Tyr Val Gln
    50                  55                  60
Gly Asn Ala Lys Phe Val Ser Pro Val Arg Arg Ile Val Val His Glu
65                  70                  75                  80
Tyr Tyr Asn Ser Gln Thr Phe Asp Tyr Asp Ile Ala Leu Leu Gln Leu
                85                  90                  95
Ser Ile Ala Trp Pro Glu Thr Leu Lys Gln Leu Ile Gln Pro Ile Cys
            100                 105                 110
Ile Pro Pro Thr Gly Gln Arg Val Arg Ser Gly Glu Lys Cys Trp Val
        115                 120                 125
Thr Gly Trp Gly Arg Arg His Glu Ala Asp Asn Lys Gly Ser Leu Val
    130                 135                 140
Leu Gln Gln Ala Glu Val Glu Leu Ile Asp Gln Thr Leu Cys Val Ser
145                 150                 155                 160
Thr Tyr Gly Ile Ile Thr Ser Arg Met Leu Cys Ala Gly Ile Met Ser
                165                 170                 175
Gly Lys Arg Asp Ala Cys Lys Gly Asp Ser Gly Gly Pro Leu Ser Cys
            180                 185                 190
Arg Arg Lys Ser Asp Gly Lys Trp Ile Leu Thr Gly Ile Val Ser Trp
        195                 200                 205
Gly His Gly Ser Gly Arg Pro Asn Phe Pro Gly Val Tyr Thr Arg Val
    210                 215                 220
Ser Asn Phe Val Pro Trp Ile His Lys Tyr Val Pro Ser Leu
225                 230                 235
```

<210> 75
<211> 235
<212> PRT
<213> Homo sapiens

<400> 75

```
Arg Ile Val Gln Gly Arg Glu Thr Ala Met Glu Gly Glu Trp Pro Trp
 1               5                  10                  15
Gln Ala Ser Leu Gln Leu Ile Gly Ser Gly His Gln Cys Gly Ala Ser
            20                  25                  30
```

```
Leu Ile Ser Asn Thr Trp Leu Leu Thr Ala Ala His Cys Phe Trp Lys
        35                  40                  45
Asn Lys Asp Pro Thr Gln Trp Ile Ala Thr Phe Gly Ala Thr Ile Thr
    50                  55                  60
Pro Pro Ala Val Lys Arg Asn Val Arg Lys Ile Ile Leu His Glu Asn
65                  70                  75                  80
Tyr His Arg Glu Thr Asn Glu Asn Asp Ile Ala Leu Val Gln Leu Ser
                85                  90                  95
Thr Gly Val Glu Phe Ser Asn Ile Val Gln Arg Val Cys Leu Pro Asp
            100                 105                 110
Ser Ser Ile Lys Leu Pro Pro Lys Thr Ser Val Phe Val Thr Gly Phe
        115                 120                 125
Gly Ser Ile Val Asp Asp Gly Pro Ile Gln Asn Thr Leu Arg Gln Ala
    130                 135                 140
Arg Val Glu Thr Ile Ser Thr Asp Val Cys Asn Arg Lys Asp Val Tyr
145                 150                 155                 160
Asp Gly Leu Ile Thr Pro Gly Met Leu Cys Ala Gly Phe Met Glu Gly
                165                 170                 175
Lys Ile Asp Ala Cys Lys Gly Asp Ser Gly Gly Pro Leu Val Tyr Asp
            180                 185                 190
Asn His Asp Ile Trp Tyr Ile Val Gly Ile Val Ser Trp Gly Gln Ser
        195                 200                 205
Cys Ala Leu Pro Lys Lys Pro Gly Val Tyr Thr Arg Val Thr Lys Tyr
    210                 215                 220
Arg Asp Trp Ile Ala Ser Lys Thr Gly Met Asn
225                 230                 235
```

<210> 76
<211> 230
<212> PRT
<213> Homo sapiens

<400> 76

```
Arg Ile Val Gly Gly Ser Ala Ala Pro Pro Gly Ala Trp Pro Trp Leu
 1               5                  10                  15
Val Arg Leu Gln Leu Gly Gly Gln Pro Leu Cys Gly Gly Val Leu Val
            20                  25                  30
Ala Ala Ser Trp Val Leu Thr Ala Ala His Cys Phe Leu Leu Trp Thr
        35                  40                  45
Val Thr Leu Ala Glu Gly Ser Arg Gly Glu Gln Ala Glu Glu Val Pro
    50                  55                  60
Val Asn Arg Ile Leu Pro His Pro Lys Phe Asp Pro Arg Thr Phe His
65                  70                  75                  80
Asn Asp Leu Ala Leu Val Gln Leu Trp Thr Pro Val Ser Pro Gly Gly
                85                  90                  95
Ser Ala Arg Pro Val Cys Leu Pro Gln Glu Pro Gln Glu Pro Pro Ala
            100                 105                 110
Gly Thr Ala Cys Ala Ile Ala Gly Trp Gly Ala Leu Phe Glu Asp Gly
        115                 120                 125
Pro Glu Ala Glu Ala Val Arg Glu Ala Arg Val Pro Leu Leu Ser Thr
    130                 135                 140
Asp Thr Cys Arg Arg Ala Leu Gly Pro Gly Leu Arg Pro Ser Thr Met
145                 150                 155                 160
Leu Cys Ala Gly Tyr Leu Ala Gly Gly Val Asp Ser Cys Gln Gly Asp
                165                 170                 175
Ser Gly Gly Pro Leu Thr Cys Ser Glu Pro Gly Pro Arg Pro Arg Glu
            180                 185                 190
Val Leu Phe Gly Val Thr Ser Trp Gly Asp Gly Cys Gly Glu Pro Gly
        195                 200                 205
Lys Pro Gly Val Tyr Thr Arg Val Ala Val Phe Lys Asp Trp Leu Gln
    210                 215                 220
```

Glu Gln Met Ser Gly Glu
225 230

<210> 77
<211> 233
<212> PRT
<213> Homo sapiens

<400> 77
Arg Ile Ala Ser Gly Val Ile Ala Pro Lys Ala Ala Trp Pro Trp Gln
1 5 10 15
Ala Ser Leu Gln Tyr Asp Asn Ile His Gln Cys Gly Ala Thr Leu Ile
20 25 30
Ser Asn Thr Trp Leu Val Thr Ala Ala His Cys Phe Gln Lys Tyr Lys
35 40 45
Asn Pro His Gln Trp Thr Val Ser Phe Gly Thr Lys Ile Asn Pro Pro
50 55 60
Leu Met Lys Arg Asn Val Arg Arg Phe Ile Ile His Glu Lys Tyr Arg
65 70 75 80
Ser Ala Ala Arg Glu Tyr Asp Ile Ala Val Val Gln Val Ser Ser Arg
85 90 95
Val Thr Phe Ser Asp Asp Ile Arg Gln Ile Cys Leu Pro Glu Ala Ser
100 105 110
Ala Ser Phe Gln Pro Asn Leu Thr Val His Ile Thr Gly Phe Gly Ala
115 120 125
Leu Tyr Tyr Gly Gly Glu Ser Gln Asn Asp Leu Arg Glu Ala Arg Val
130 135 140
Lys Ile Ile Ser Asp Asp Val Cys Lys Gln Pro Gln Val Tyr Gly Asn
145 150 155 160
Asp Ile Lys Pro Gly Met Phe Cys Ala Gly Tyr Met Glu Gly Ile Tyr
165 170 175
Asp Ala Cys Arg Gly Asp Ser Gly Gly Pro Leu Val Thr Arg Asp Leu
180 185 190
Lys Asp Thr Trp Tyr Leu Ile Gly Ile Val Ser Trp Gly Asp Asn Cys
195 200 205
Gly Gln Lys Asp Lys Pro Gly Val Tyr Thr Gln Val Thr Tyr Tyr Arg
210 215 220
Asn Trp Ile Ala Ser Lys Thr Gly Ile
225 230

<210> 78
<211> 247
<212> PRT
<213> Homo sapiens

<400> 78
Arg Ile Ile Gly Gly Thr Glu Ala Gln Ala Gly Ala Trp Pro Trp Val
1 5 10 15
Val Ser Leu Gln Ile Lys Tyr Gly Arg Val Leu Val His Val Cys Gly
20 25 30
Gly Thr Leu Val Arg Glu Arg Trp Val Leu Thr Ala Ala His Cys Thr
35 40 45
Lys Asp Ala Ser Asp Pro Leu Met Trp Thr Ala Val Ile Gly Thr Asn
50 55 60
Asn Ile His Gly Arg Tyr Pro His Thr Lys Lys Ile Lys Ile Lys Ala
65 70 75 80
Ile Ile Ile His Pro Asn Phe Ile Leu Glu Ser Tyr Val Asn Asp Ile
85 90 95
Ala Leu Phe His Leu Lys Lys Ala Val Arg Tyr Asn Asp Tyr Ile Gln
100 105 110

```
Pro Ile Cys Leu Pro Phe Asp Val Phe Gln Ile Leu Asp Gly Asn Thr
        115                 120                 125
Lys Cys Phe Ile Ser Gly Trp Gly Arg Thr Lys Glu Glu Gly Asn Tyr
    130                 135                 140
Gly Asn Ala Thr Asn Ile Leu Gln Asp Ala Glu Val His Tyr Ile Ser
145                 150                 155                 160
Arg Glu Met Cys Asn Ser Glu Arg Ser Tyr Gly Gly Ile Ile Pro Asn
                165                 170                 175
Thr Ser Phe Cys Ala Gly Asp Glu Asp Gly Ala Phe Asp Thr Cys Arg
            180                 185                 190
Gly Asp Ser Gly Gly Pro Leu Met Cys Tyr Leu Pro Glu Tyr Lys Arg
        195                 200                 205
Phe Phe Val Met Gly Ile Thr Ser Tyr Gly His Gly Cys Gly Arg Arg
    210                 215                 220
Gly Phe Pro Gly Val Tyr Ile Gly Pro Ser Phe Tyr Gln Lys Trp Leu
225                 230                 235                 240
Thr Glu His Phe Phe His Ala
                245
```

<210> 79
<211> 253
<212> PRT
<213> Homo sapiens

<400> 79

```
Arg Ile Ser Ser Trp Arg Asn Ser Thr Val Thr Gly His Pro Trp Gln
 1               5                  10                  15
Val Ser Leu Lys Ser Asp Glu His His Phe Cys Gly Gly Ser Leu Ile
            20                  25                  30
Gln Glu Asp Arg Val Val Thr Ala Ala His Cys Leu Asp Ser Leu Ser
        35                  40                  45
Glu Lys Gln Leu Lys Asn Ile Thr Val Thr Ser Gly Glu Tyr Ser Leu
    50                  55                  60
Phe Gln Lys Asp Lys Gln Glu Gln Asn Ile Pro Val Ser Lys Ile Ile
65                  70                  75                  80
Thr His Pro Glu Tyr Asn Ser Arg Glu Tyr Met Ser Pro Asp Ile Ala
                85                  90                  95
Leu Leu Tyr Leu Lys His Lys Val Lys Phe Gly Asn Ala Val Gln Pro
            100                 105                 110
Ile Cys Leu Pro Asp Ser Asp Asp Lys Val Glu Pro Gly Ile Leu Cys
        115                 120                 125
Leu Ser Ser Gly Trp Gly Lys Ile Ser Lys Thr Ser Glu Tyr Ser Asn
    130                 135                 140
Val Leu Gln Glu Met Glu Leu Pro Ile Met Asp Asp Arg Ala Cys Asn
145                 150                 155                 160
Thr Val Leu Lys Ser Met Asn Leu Pro Pro Leu Gly Arg Thr Met Leu
                165                 170                 175
Cys Ala Gly Phe Pro Asp Trp Gly Met Asp Ala Cys Gln Gly Asp Ser
            180                 185                 190
Gly Gly Pro Leu Val Cys Arg Arg Gly Gly Gly Ile Trp Ile Leu Ala
        195                 200                 205
Gly Ile Thr Ser Trp Val Ala Gly Cys Ala Gly Gly Ser Val Pro Val
    210                 215                 220
Arg Asn Asn His Val Lys Ala Ser Leu Gly Ile Phe Ser Lys Val Ser
225                 230                 235                 240
Glu Leu Met Asp Phe Ile Thr Gln Asn Leu Phe Thr Gly
                245                 250
```

<210> 80
<211> 241

<212> PRT
<213> Homo sapiens

<400> 80
Glu Ile Trp Ser Gly Glu Gln Gly Gln Asn Asp Met Val Trp Leu Ser
1 5 10 15
Ser Leu Lys Met Ser Gly Gln His Tyr Cys Gly Ala Ser Leu Ile Ser
20 25 30
Glu Arg His Leu Val Thr Ala Ala His Cys Phe Lys Val Thr Lys Asn
35 40 45
Pro Lys Asn Tyr Thr Val Ser Phe Gly Thr Lys Val Thr Leu Pro Tyr
50 55 60
Met Gln His Asp Val Gln Gln Ile Ile Ile His Glu Asp Tyr Ile Gln
65 70 75 80
Asp Glu His His Asp Asp Ile Ala Leu Ile Leu Leu Thr Lys Lys Val
85 90 95
Leu Phe Lys Asn Asp Val His Arg Val Cys Leu Pro Glu Ala Thr Gln
100 105 110
Ile Phe Pro Pro Gly Glu Gly Val Val Val Thr Gly Trp Gly Arg Leu
115 120 125
Ser Phe Asn Gly Lys Ile Ser Glu Asn Leu Thr Tyr His Lys Ala Ser
130 135 140
Val Lys Ile Thr Asp Thr Asn Thr Cys Asn Ala Lys Glu Ala Tyr Arg
145 150 155 160
Ser Met Val Gln Asp Arg Val Leu Cys Ala Gly Tyr Met Glu Gly Asn
165 170 175
Ile Asp Ala Cys Gln Gly Asp Ser Gly Gly Pro Leu Val His Pro Asn
180 185 190
Ser Leu Asn Ile Trp Tyr Ile Trp Tyr Leu Val Gly Val Val Ser Trp
195 200 205
Gly Arg Asn Glu Cys Gly Ala Ile Asn Ser Pro Gly Val Tyr Thr Gln
210 215 220
Thr Asp Val Phe Phe Phe Leu Lys Trp Ile Lys Ser Thr Ile Ala Leu
225 230 235 240
Lys

<210> 81
<211> 231
<212> PRT
<213> Homo sapiens

<400> 81
Arg Ile Val Ser Met Glu Ser Lys Lys Gly Lys Val Gln Trp Leu Val
1 5 10 15
Val Leu Phe Gly Ser Ser Ser Ile Gln Gly Ser Arg Lys Asp Lys Ala
20 25 30
Ile Lys Thr Trp Thr Thr Phe Ser Tyr Thr Val Trp Leu Gly Ser Ile
35 40 45
Thr Val Gly Asp Ser Arg Lys Arg Val Lys Tyr Tyr Val Ser Lys Ile
50 55 60
Val Ile His Pro Lys Tyr Gln Asp Thr Thr Ala Asp Val Ala Leu Leu
65 70 75 80
Lys Leu Ser Ser Gln Val Thr Phe Thr Ser Ala Ile Leu Pro Ile Cys
85 90 95
Leu Pro Ser Val Thr Lys Gln Leu Ala Ile Pro Pro Phe Cys Trp Val
100 105 110
Thr Gly Trp Gly Lys Val Lys Glu Ser Ser Asp Arg Asp Tyr His Ser
115 120 125
Ala Leu Gln Glu Ala Glu Val Pro Ile Ile Asp Arg Gln Ala Cys Glu
130 135 140

```
Gln Leu Tyr Asn Pro Ile Gly Ile Phe Leu Pro Ala Leu Glu Pro Val
145                 150                 155                 160
Ile Lys Glu Asp Lys Ile Cys Ala Gly Asp Thr Gln Asn Met Lys Asp
                165                 170                 175
Ser Cys Lys Gly Asp Ser Gly Gly Pro Leu Ser Cys His Ile Asp Gly
            180                 185                 190
Val Trp Ile Gln Thr Gly Val Val Ser Trp Gly Leu Glu Cys Gly Lys
        195                 200                 205
Ser Leu Pro Gly Val Tyr Thr Asn Val Ile Tyr Tyr Gln Lys Trp Ile
    210                 215                 220
Asn Ala Thr Ile Ser Arg Ala
225                 230
```

<210> 82
<211> 223
<212> PRT
<213> Homo sapiens

```
<400> 82
Leu Ala Phe Asn Pro Asp Tyr Thr Val Ser Ser Thr Pro Pro Tyr Leu
 1               5                  10                  15
Val Tyr Leu Lys Ser Asp Tyr Leu Pro Cys Ala Gly Val Leu Ile His
            20                  25                  30
Pro Leu Trp Val Ile Thr Ala Ala His Cys Asn Leu Pro Lys Leu Arg
        35                  40                  45
Val Ile Leu Gly Val Thr Ile Pro Ala Asp Ser Asn Glu Lys His Leu
    50                  55                  60
Gln Val Ile Gly Tyr Glu Lys Met Ile His His Pro His Phe Ser Val
65                  70                  75                  80
Thr Ser Ile Asp His Asp Ile Met Leu Ile Lys Leu Lys Thr Glu Ala
                85                  90                  95
Glu Leu Asn Asp Tyr Val Lys Leu Ala Asn Leu Pro Tyr Gln Thr Ile
            100                 105                 110
Ser Glu Asn Thr Met Cys Ser Val Ser Thr Trp Ser Tyr Asn Val Tyr
        115                 120                 125
Lys Glu Pro Asp Ser Leu Gln Thr Val Asn Ile Ser Val Ile Ser Lys
    130                 135                 140
Pro Gln Cys Arg Asp Ala Tyr Lys Thr Tyr Asn Ile Thr Glu Asn Met
145                 150                 155                 160
Leu Cys Val Gly Ile Val Pro Gly Arg Arg Gln Pro Cys Lys Glu Val
                165                 170                 175
Ser Ala Ala Pro Ala Ile Cys Asn Gly Met Leu Gln Gly Ile Leu Ser
            180                 185                 190
Phe Ala Asp Gly Cys Val Leu Arg Ala Asp Val Gly Ile Tyr Ala Lys
        195                 200                 205
Ile Phe Tyr Tyr Ile Pro Trp Ile Glu Asn Val Ile Gln Asn Asn
    210                 215                 220
```

<210> 83
<211> 223
<212> PRT
<213> Homo sapiens

```
<400> 83
Arg Trp Ala Ala Gly Val Arg Val Pro Ala Gln His Ser Glu Glu Pro
 1               5                  10                  15
Pro His Asn Arg Ser Thr Asn Pro Ser Asp Tyr Arg Ile Leu Leu Gly
            20                  25                  30
Tyr Asp Gln Gln Ser His Pro Thr Glu His Ser Lys Gln Met Thr Val
        35                  40                  45
```

```
Asn Lys Ile Met Val His Ala Asp Tyr Asn Glu Leu His Arg Met Gly
    50                  55                  60
Ser Asp Ile Thr Leu Leu Gln Leu His His His Val Glu Phe Ser Ser
65                  70                  75                  80
His Ile Leu Pro Ala Cys Leu Pro Glu Pro Thr Thr Trp Leu Ala Pro
                85                  90                  95
Asp Ser Ser Cys Trp Ile Ser Gly Trp Gly Met Val Thr Glu Asp Val
            100                 105                 110
Phe Leu Pro Glu Pro Phe Gln Leu Gln Glu Ala Glu Val Gly Val Met
        115                 120                 125
Asp Asn Thr Val Cys Gly Ser Phe Phe Gln Pro Gln Tyr Pro Gly Gln
    130                 135                 140
Pro Ser Ser Ser Asp Tyr Thr Ile His Glu Asp Met Leu Cys Ala Gly
145                 150                 155                 160
Asp Leu Ile Thr Gly Lys Ala Ile Cys Arg Arg Asp Ser Arg Gly Pro
                165                 170                 175
Leu Val Cys Pro Leu Asn Gly Thr Trp Phe Leu Met Gly Leu Ser Ser
            180                 185                 190
Trp Ser Leu Asp Cys Cys Ser Pro Val Gly Pro Arg Val Phe Thr Arg
        195                 200                 205
Leu Pro Tyr Phe Thr Asn Trp Ile Ser Gln Lys Lys Arg Glu Ser
    210                 215                 220
```

<210> 84
<211> 203
<212> PRT
<213> Homo sapiens

<400> 84

```
Arg Val Val Ser Gly Tyr Phe Ser Ala Asn Met Val Ser Thr Pro Trp
 1               5                  10                  15
Arg Thr Gly Ile Leu His Phe Asn His Cys Ile His Asp Leu Ser Gln
            20                  25                  30
Thr Val Leu Gly Asp His Leu Val Lys Phe His His Thr Ile Lys Ile
        35                  40                  45
Ile Cys His Ile Leu Asp His Ala Val Ala Leu Leu Phe Leu Gln Ile
    50                  55                  60
Ser Ser Ile Trp Asn Gly Asn Ile Tyr Pro Ile Pro Leu Pro Ala Phe
65                  70                  75                  80
Val Ser Tyr Lys Asn Ala Ser Ile Cys Arg Ile Met Leu Trp Gly His
                85                  90                  95
Ala Gly Asp Met Leu Phe Pro Met Asn Phe Pro Leu Cys Ala Arg Val
            100                 105                 110
Asp Arg Gln Gln Gly Glu Gln Cys Glu His Thr Glu Phe Gly Tyr Gln
        115                 120                 125
Pro Glu Thr Ile Lys Asn Asp Met Leu Cys Ala Gly Phe Glu Glu Gly
    130                 135                 140
Lys Lys Asp Ala Cys Lys Gly Asp Ser Gly Gly Pro Leu Val Cys Leu
145                 150                 155                 160
Val Gly Gln Ser Trp Leu Gln Ala Gly Val Ile Ser Trp Gly Glu Gly
                165                 170                 175
Cys Ala Arg Gln Asn Arg Pro Gly Val Tyr Ile Arg Val Thr Ala His
            180                 185                 190
His Asn Trp Ile His Arg Ile Ile Pro Lys Leu
        195                 200
```

<210> 85
<211> 235
<212> PRT
<213> Homo sapiens

<400> 85
His Ile Ile Asn Gly Lys Arg Gln Ile Ala Phe Pro Arg Arg Pro Gly
1 5 10 15
Thr Arg Glu Gly Cys Pro Leu Leu Leu Phe Leu Ser Asn Ala His Cys
20 25 30
Thr Pro Pro Trp Ala Thr Glu Gln Asp Ser Asn Ser Lys Lys Lys Lys
35 40 45
Lys Lys Glu Thr Glu Lys Thr Ile Pro Lys Ala Thr Val Ile Lys Thr
50 55 60
Asp Gly His Tyr Lys Glu Asn Lys Asn Arg Lys His Gln Val Leu Ala
65 70 75 80
Lys Met Trp Arg Asn Trp Asn Leu Tyr Ala Leu Leu Val Phe Cys Lys
85 90 95
Ile Lys His Arg Ile Thr Glu Pro Gly Arg Val Ala His Ala Cys Asn
100 105 110
Pro Ser Thr Leu Gly Gly Arg Gly Gly Trp Ile Thr Arg Trp Gly Ser
115 120 125
His Tyr Val Ala Gln Ala Gly Glu Thr Ser Asp Glu Leu Gln Glu Met
130 135 140
Gln Leu Pro Leu Ile Leu Glu Pro Trp Cys His Leu Leu Tyr Gly His
145 150 155 160
Met Ser Tyr Ile Met Pro Asp Met Leu Cys Ala Gly Asp Ile Leu Asn
165 170 175
Ala Lys Thr Val Cys Glu Gly Asp Ser Gly Gly Pro Leu Val Cys Glu
180 185 190
Phe Asn Arg Ser Trp Leu Gln Ile Gly Ile Val Ser Trp Gly Arg Gly
195 200 205
Cys Ser Asn Pro Leu Tyr Pro Gly Val Tyr Ala Ser Val Ser Tyr Phe
210 215 220
Ser Lys Trp Ile Cys Asp Asn Ile Glu Ile Thr
225 230 235

<210> 86
<211> 228
<212> PRT
<213> Homo sapiens

<400> 86
Arg Val Ser Gly Gly Arg Asp Ser Val Pro Ser Leu Val Pro Ser Thr
1 5 10 15
Asn Ala Tyr Asn Arg Lys Arg Pro Glu Asn Pro His Met Cys Gly Gly
20 25 30
Phe Leu Ala Ser Asn Ile Glu His Leu Leu Cys Ala Arg His Arg Ile
35 40 45
Gln Lys Ser Met Thr Ser Ala His Arg Ser Lys Val Arg Arg Leu Glu
50 55 60
Ser His Trp Tyr Lys Gly Lys Arg Lys Thr Arg Ser Lys Glu Lys Arg
65 70 75 80
Lys Ile Phe Gly Lys Tyr Thr Ser Asn Ile Asn Tyr Asp Ile Ser Leu
85 90 95
Leu Gly Leu Ala Ser Pro Ala Val Ile Thr Asp Lys Val Ile Pro Ala
100 105 110
Cys Leu Pro Ser Pro Asn Tyr Val Val Ala Asp Gln Thr Glu Cys Tyr
115 120 125
Ile Thr Asp Trp Gly Glu Thr Gln Gly Thr Phe Gly Ala Gly Phe Leu
130 135 140
Lys Glu Ala Gln Leu Pro Val Ile Glu Asn Glu Val Cys Asn Arg Tyr
145 150 155 160
Glu Phe Leu Asn Gly Arg Val Lys Ser Thr Glu Leu Cys Ala Gly His
165 170 175

Leu Ala Gly Gly Ile Asp Ser Cys Lys Val Arg Lys Asp Gln Glu Thr
180 185 190
Lys Val Ser Leu Phe Gly Ile Gly Cys Gly Asp Trp Val Arg Ser Pro
195 200 205
His Phe Tyr Thr Tyr Ile His Thr Tyr Thr Pro Ser Ile Gln Glu Asn
210 215 220
Ile Lys Glu Asn
225

<210> 87
<211> 252
<212> PRT
<213> Homo sapiens

<400> 87
Arg Lys Leu Gly Ile Leu Asn His Gln Val Leu Phe Trp Tyr Asn Leu
1 5 10 15
Ser Leu Leu Leu His Phe Ile Gly Tyr Lys Ser Tyr Ser Glu Pro Leu
20 25 30
Ala Leu Phe Gly Glu Asp Asp Asp Met Asp Pro Arg Pro Ser Arg Ser
35 40 45
Tyr Gln Val Ala Asn Gly Ile Ala Val Leu Pro Val Ser Gly Thr Leu
50 55 60
Val Ser Lys Thr Arg Ala Leu Gln Pro Tyr Ser Gly Met Thr Gly Tyr
65 70 75 80
Asn Gly Ile Ile Ala Arg Leu Gln Gln Ala Ile Ser Asp Pro Gly Val
85 90 95
Asp Gly Ile Leu Leu Asp Met Asp Thr Pro Gly Gly Met Val Ser Gly
100 105 110
Ala Phe Asp Cys Ala Asp Ile Ile Ala Arg Met Arg Asp Ile Lys Pro
115 120 125
Ile Trp Ala Leu Ala Asn Asp Met Asn Cys Ser Ala Gly Gln Leu Ile
130 135 140
Ala Ser Ser Ala Ser Arg Arg Leu Val Thr Gln Thr Ala Arg Thr Gly
145 150 155 160
Ser Ile Gly Val Met Met Ala His Ser Asn Tyr Gly Ala Ala Leu Lys
165 170 175
Thr Asn Gly Gly His Met His Thr Tyr Val Tyr Cys Ser Thr Ile His
180 185 190
Asn Ser Lys Asp Leu Lys Pro Thr Gln Met Pro Ile Asn Asn Arg Leu
195 200 205
Asp Lys Glu Asn Val Ala His Ile His His Gly Ile Leu Cys Ser His
210 215 220
Lys Lys Asp Glu Phe Met Ser Phe Ala Gly Thr Trp Met Lys Leu Glu
225 230 235 240
Thr Ile Ile Leu Ser Lys Leu Thr Gln Glu Gln Lys
245 250

<210> 88
<211> 299
<212> PRT
<213> Homo sapiens

<400> 88
Met Leu Gly Val Leu Leu Gln Ile Trp Arg Gly Ser Trp Lys Lys Gln
1 5 10 15
Thr Gln Ala Gln Gly Arg Arg Glu Arg Ser Arg Gln Ala Ala Gly Ala
20 25 30
Val Ser Ala Gly Gly Arg Arg Ala Leu Leu Leu Tyr Leu Arg Ala Glu
35 40 45

```
Leu Glu Asp Lys Leu Ala Cys Val Asp Ser Arg Leu Arg Leu Val Met
    50                  55                  60
Arg Gly Leu Val Leu Gly Arg Ala Ser Gly Ser Ser Val Arg Pro Lys
65                  70                  75                  80
Leu Pro Lys Asp Val Arg Ala Asp Phe Gln Thr Arg Ile Asp Ala Thr
                85                  90                  95
Arg Gln Met Phe Ala Glu Lys Val Ser Ala Tyr Thr Gly Met Ser Val
            100                 105                 110
Gln Asp Val Leu Asp Thr Glu Ala Ala Val Phe Ser Gly Gln Glu Ser
        115                 120                 125
Leu Asp Asn Gly Leu Ala Asp Glu Leu Val Asn Asn Thr Asp Ala Leu
    130                 135                 140
Gly Val Met Arg Glu Ala Leu Asp Arg Arg Lys Lys Thr Thr Leu Gly
145                 150                 155                 160
Gly Thr Met Pro Ser Pro Ser Ala Ser Ala Val Thr Thr Lys Pro Val
                165                 170                 175
Asp Gln Ala Ala Thr Gln Thr Thr Ala Ser Ala Glu Gln Ala Thr Thr
            180                 185                 190
Val Asp Thr Thr Ile Ala Ser Val Ala Ala Pro Val Asp Val Ser Ala
        195                 200                 205
Gln Val Thr Ala Ala Val Ala Ala Glu Asn Ser Arg Ile Met Gly Ile
    210                 215                 220
Leu Asn Cys Asp Glu Ala Lys Gly Arg Glu Ser Gln Ala Arg Ala Leu
225                 230                 235                 240
Ala Glu Thr Pro Gly Met Thr Val Glu Ser Ala Gln Arg Ile Leu Ala
                245                 250                 255
Ala Ala Pro Gln Ser Ala Gln Met Arg Thr Asp Thr Ala Leu Asp Arg
            260                 265                 270
Leu Met Glu Thr Ala Pro Gly Ala Leu Gln Ala Gly Ser Ala Ser Ser
        275                 280                 285
Asp Ala Ala Asp Asp Leu Leu Asn Thr Pro Val
    290                 295
```

<210> 89
<211> 463
<212> PRT
<213> Homo sapiens

```
<400> 89
Thr Asp Pro Trp Phe Ser Lys Gln Trp Tyr Met Asn Ser Glu Ala Gln
 1               5                  10                  15
Pro Asp Leu Ser Ile Leu Gln Ala Trp Ser Gln Gly Leu Ser Gly Gln
            20                  25                  30
Gly Ile Val Val Ser Val Leu Asp Asp Gly Ile Glu Lys Asp His Pro
        35                  40                  45
Asp Leu Trp Ala Asn Tyr Asp Pro Leu Ala Ser Tyr Asp Phe Asn Asp
    50                  55                  60
Tyr Asp Pro Asp Pro Gln Pro Arg Tyr Thr Pro Ser Lys Glu Asn Arg
65                  70                  75                  80
His Gly Thr Arg Cys Ala Gly Glu Val Ala Ala Met Ala Asn Asn Gly
                85                  90                  95
Phe Cys Gly Val Gly Val Ala Phe Asn Ala Arg Ile Gly Gly Val Arg
            100                 105                 110
Met Leu Asp Gly Thr Ile Thr Asp Val Ile Glu Ala Gln Ser Leu Ser
        115                 120                 125
Leu Gln Pro Gln His Ile His Ile Tyr Ser Ala Ser Trp Gly Pro Glu
    130                 135                 140
Asp Asp Gly Arg Thr Val Asp Gly Pro Gly Ile Leu Thr Arg Glu Ala
145                 150                 155                 160
Phe Arg Arg Gly Val Thr Lys Gly Arg Gly Gly Leu Gly Thr Leu Phe
                165                 170                 175
```

```
Ile Trp Ala Ser Gly Asn Gly Gly Leu His Tyr Asp Asn Cys Asn Cys
            180                 185                 190
Asp Gly Tyr Thr Asn Ser Ile His Thr Leu Ser Val Gly Ser Thr Thr
        195                 200                 205
Gln Gln Gly Arg Val Pro Trp Tyr Ser Glu Ala Cys Ala Ser Thr Leu
    210                 215                 220
Thr Thr Thr Tyr Ser Ser Gly Val Ala Thr Asp Pro Gln Ile Val Thr
225                 230                 235                 240
Thr Asp Leu His His Gly Cys Thr Asp Gln His Thr Gly Thr Ser Ala
                245                 250                 255
Ser Ala Pro Leu Ala Ala Gly Met Ile Ala Leu Ala Leu Glu Ala Asn
            260                 265                 270
Pro Phe Leu Thr Trp Arg Asp Met Gln His Leu Val Val Arg Ala Ser
        275                 280                 285
Lys Pro Ala His Leu Gln Ala Glu Asp Trp Arg Thr Asn Gly Val Gly
    290                 295                 300
Arg Gln Val Ser His His Tyr Gly Tyr Gly Leu Leu Asp Ala Gly Leu
305                 310                 315                 320
Leu Val Asp Thr Ala Arg Thr Trp Leu Pro Thr Gln Pro Gln Arg Lys
                325                 330                 335
Cys Ala Val Arg Val Gln Ser Arg Pro Thr Pro Ile Leu Pro Leu Ile
            340                 345                 350
Tyr Ile Arg Glu Asn Val Ser Ala Cys Ala Gly Leu His Asn Ser Ile
        355                 360                 365
Arg Ser Leu Glu His Val Gln Ala Gln Leu Thr Leu Ser Tyr Ser Arg
    370                 375                 380
Arg Gly Asp Leu Glu Ile Ser Leu Thr Ser Pro Met Gly Thr Arg Ser
385                 390                 395                 400
Thr Leu Val Ala Ile Arg Pro Leu Asp Val Ser Thr Glu Gly Tyr Asn
                405                 410                 415
Asn Trp Val Phe Met Ser Thr His Phe Trp Asp Glu Asn Pro Gln Gly
            420                 425                 430
Val Trp Thr Leu Gly Leu Glu Asn Lys Gly Tyr Tyr Phe Asn Thr Gly
        435                 440                 445
Glu Gly Gly Ala Gly Leu Trp Trp Ala Gly Leu Gly Ser Pro Thr
    450                 455                 460
```

<210> 90
<211> 225
<212> PRT
<213> Homo sapiens

<400> 90

```
Met Ala Ser Arg Tyr Asp Arg Ala Ile Thr Val Phe Ser Pro Asp Gly
 1               5                  10                  15
His Leu Phe Gln Val Glu Tyr Ala Gln Glu Ala Val Lys Lys Gly Ser
            20                  25                  30
Thr Ala Val Gly Ile Arg Gly Thr Asn Ile Val Val Leu Gly Val Glu
        35                  40                  45
Lys Lys Ser Val Ala Lys Leu Gln Asp Glu Arg Thr Val Arg Lys Ile
    50                  55                  60
Cys Ala Leu Asp Asp His Val Cys Met Ala Phe Ala Gly Leu Thr Ala
65                  70                  75                  80
Asp Ala Arg Val Val Ile Asn Arg Ala Arg Val Glu Cys Gln Ser His
                85                  90                  95
Lys Leu Thr Val Glu Asp Pro Val Thr Val Glu Tyr Ile Thr Arg Phe
            100                 105                 110
Ile Ala Thr Leu Lys Gln Ile Asn Thr Lys Ser Tyr Leu Lys Phe Ser
        115                 120                 125
Arg Glu Val Pro Phe Leu Phe Cys Phe Leu Phe Phe Ser Trp Asp Tyr
    130                 135                 140
```

```
Arg His Met Pro Pro His Leu Ala Asn Phe Phe Ala Gly Tyr Lys Ile
145                 150                 155                 160
Asn Lys Gln Lys Phe Ala Ala Phe Leu Tyr Ala Asn Asn Glu Gln Ser
                165                 170                 175
Glu Lys Glu Ile Lys Lys Val Ile Pro Phe Met Ile Ala Thr Asn Lys
            180                 185                 190
Ile Lys Cys Ile Glu Ile Asn Leu Thr Lys Glu Val Lys Asp Phe His
        195                 200                 205
Asn Glu Asn Tyr Lys Thr Leu Met Gln Glu Thr Glu Ala Asp Thr Lys
    210                 215                 220
Lys
225
```

<210> 91
<211> 228
<212> PRT
<213> Homo sapiens

<400> 91

```
Ser Lys Gly Gly Ile Ser Val Gly Leu Cys Val Arg Asp Gly Val Val
 1               5                  10                  15
Val Val Ser Arg Asp Thr Asn Ser Pro His Arg Val Thr Pro Leu Leu
            20                  25                  30
Asn Glu Leu Met Cys Leu Arg Cys Ser Gly Leu Ala Ala Ala Ala Lys
        35                  40                  45
Met Val Ala Ala Phe Ile Ser Leu Arg Arg Ser Ala Glu Ile Asn Lys
    50                  55                  60
Tyr Val Ile Tyr Pro Arg Asp Val Cys Thr Pro Tyr Ile Val Asn Arg
65                  70                  75                  80
Met Ser Leu Ile Lys Ile Lys Tyr Thr Gln Ser Asn Gly Arg Arg Pro
                85                  90                  95
Phe Gly Ile Ser Ala Leu Ile Val Gly Phe Asp Asp Asp Gly Ile Ser
            100                 105                 110
Arg Leu Tyr Gln Thr Asp Pro Ser Gly Thr Tyr His Ala Trp Lys Ala
        115                 120                 125
Asn Ala Ile Gly Arg Ser Ala Lys Thr Val Arg Glu Phe Leu Glu Lys
    130                 135                 140
Asn Tyr Thr Glu Asp Ala Ile Ala Ser Asp Ser Glu Ala Ile Lys Leu
145                 150                 155                 160
Ala Ile Lys Ala Leu Leu Glu Val Val Gln Ser Gly Gly Lys Asn Ile
                165                 170                 175
Glu Leu Ala Ile Ile Arg Arg Asn Gln Pro Leu Lys Lys Lys Glu Glu
            180                 185                 190
Glu Glu Glu Arg Arg Lys Lys Lys Glu Glu Glu Gly Gly Glu Glu Glu
        195                 200                 205
Glu Glu Glu Glu Glu Glu Asp Glu Glu Glu Glu Glu Glu Val Glu Glu
    210                 215                 220
Glu Glu Glu Glu
225
```

<210> 92
<211> 1005
<212> PRT
<213> Homo sapiens

<400> 92

```
Glu Asn Gly Ser Leu Thr Trp Gln Glu Leu Leu Arg Gln Thr Gly Lys
 1               5                  10                  15
Cys Ser Ile Pro Cys Leu Ile Asp Thr Gly Ala Gln Ala Asn Ile Ile
            20                  25                  30
```

```
Thr Glu Glu Thr Val Arg Ala His Lys Leu Pro Thr Arg Pro Trp Ser
            35                  40                  45
Lys Ser Val Ile Tyr Gly Gly Val Tyr Pro Asn Lys Ile Asn Arg Lys
        50                  55                  60
Thr Ile Lys Leu Asn Ile Ser Leu Asn Gly Ile Ser Ile Lys Thr Glu
65                  70                  75                  80
Phe Leu Val Val Lys Lys Phe Ser His Pro Ala Ala Ile Ser Phe Thr
                85                  90                  95
Thr Leu Tyr Asp Asn Asn Ile Glu Ile Ser Ser Ser Lys His Thr Leu
            100                 105                 110
Ser Gln Met Asn Lys Val Ser Asn Ile Val Lys Glu Pro Glu Leu Pro
        115                 120                 125
Asp Ile Tyr Lys Glu Phe Lys Asp Ile Thr Ala Glu Thr Asn Thr Glu
    130                 135                 140
Lys Leu Pro Lys Pro Ile Lys Gly Leu Glu Phe Glu Val Glu Leu Thr
145                 150                 155                 160
Gln Glu Asn Tyr Arg Leu Pro Ile Arg Asn Tyr Pro Leu Pro Pro Gly
                165                 170                 175
Lys Met Gln Ala Met Asn Asp Glu Ile Asn Gln Gly Leu Lys Ser Gly
            180                 185                 190
Ile Ile Arg Glu Ser Lys Ala Ile Asn Ala Cys Pro Val Met Phe Val
        195                 200                 205
Pro Lys Lys Glu Gly Thr Leu Arg Met Val Val Asp Tyr Lys Pro Leu
    210                 215                 220
Asn Lys Tyr Val Lys Pro Asn Ile Tyr Pro Leu Pro Leu Ile Glu Gln
225                 230                 235                 240
Leu Leu Ala Lys Ile Gln Gly Ser Thr Ile Phe Thr Lys Leu Asp Leu
                245                 250                 255
Lys Ser Ala Tyr His Leu Ile Arg Val Arg Lys Gly Asp Glu His Lys
            260                 265                 270
Leu Ala Phe Arg Cys Pro Arg Gly Val Phe Glu Tyr Leu Val Met Pro
        275                 280                 285
Tyr Gly Ile Ser Ile Ala Pro Ala His Phe Gln Tyr Phe Ile Asn Thr
    290                 295                 300
Ile Leu Gly Glu Val Lys Glu Ser His Val Val Cys Tyr Met Asp Asn
305                 310                 315                 320
Ile Leu Ile His Ser Lys Ser Glu Ser Glu His Val Lys His Val Lys
                325                 330                 335
Asp Val Leu Gln Lys Leu Lys Asn Ala Asn Leu Ile Ile Asn Gln Ala
            340                 345                 350
Lys Cys Glu Phe His Gln Ser Gln Val Lys Phe Ile Gly Tyr His Ile
        355                 360                 365
Ser Glu Lys Gly Phe Thr Pro Cys Gln Glu Asn Ile Asp Lys Val Leu
    370                 375                 380
Gln Trp Lys Gln Pro Lys Asn Arg Lys Glu Leu Arg Gln Phe Leu Gly
385                 390                 395                 400
Ser Val Asn Tyr Leu Arg Lys Phe Ile Pro Lys Thr Ser Gln Leu Thr
                405                 410                 415
His Pro Leu Asn Asn Leu Leu Lys Lys Asp Val Arg Trp Lys Trp Thr
            420                 425                 430
Pro Thr Gln Thr Gln Ala Ile Glu Asn Ile Lys Gln Cys Leu Val Ser
        435                 440                 445
Pro Pro Val Leu Arg His Phe Asp Phe Ser Lys Lys Ile Leu Leu Glu
    450                 455                 460
Thr Asp Ala Ser Asp Val Ala Val Gly Ala Val Leu Ser Gln Lys His
465                 470                 475                 480
Asp Asp Asp Lys Tyr Tyr Pro Val Gly Tyr Tyr Ser Ala Lys Met Ser
                485                 490                 495
Lys Ala Gln Leu Asn Tyr Ser Val Ser Asp Lys Glu Met Leu Ala Ile
            500                 505                 510
Ile Lys Ser Leu Lys His Trp Arg His Tyr Leu Glu Ser Thr Ile Glu
        515                 520                 525
```

```
Pro Phe Lys Ile Leu Thr Asp His Arg Asn Leu Ile Gly Arg Ile Thr
    530                 535                 540
Asn Glu Ser Glu Pro Glu Asn Lys Arg Leu Ala Arg Trp Gln Leu Phe
545                 550                 555                 560
Leu Gln Asp Phe Asn Phe Glu Ile Asn Tyr Arg Pro Gly Ser Ala Asn
                565                 570                 575
His Ile Ala Asp Ala Leu Ser Arg Ile Val Asp Glu Thr Glu Pro Ile
            580                 585                 590
Pro Lys Asp Ser Glu Asp Asn Ser Ile Asn Phe Val Asn Gln Ile Ser
        595                 600                 605
Ile Thr Asp Asp Phe Lys Asn Gln Val Val Thr Glu Tyr Thr Asn Asp
    610                 615                 620
Thr Lys Leu Leu Asn Leu Leu Asn Asn Glu Asp Lys Arg Val Glu Glu
625                 630                 635                 640
Asn Ile Gln Leu Lys Asp Gly Leu Leu Ile Asn Ser Lys Asp Gln Ile
                645                 650                 655
Leu Leu Pro Asn Asp Thr Gln Leu Thr Arg Thr Ile Ile Lys Lys Tyr
            660                 665                 670
His Glu Glu Gly Lys Leu Ile His Pro Gly Ile Glu Leu Leu Thr Asn
        675                 680                 685
Ile Ile Leu Arg Arg Phe Thr Trp Lys Gly Ile Arg Lys Gln Ile Gln
    690                 695                 700
Glu Tyr Val Gln Asn Cys His Thr Cys Gln Ile Asn Lys Ser Arg Asn
705                 710                 715                 720
His Lys Pro Tyr Gly Pro Leu Gln Pro Ile Pro Pro Ser Glu Arg Pro
                725                 730                 735
Trp Glu Ser Leu Ser Met Asp Phe Ile Thr Ala Leu Pro Glu Ser Ser
            740                 745                 750
Gly Tyr Asn Ala Leu Phe Val Val Val Asp Arg Phe Ser Lys Met Ala
        755                 760                 765
Ile Leu Val Pro Cys Thr Lys Ser Ile Thr Ala Glu Gln Thr Ala Arg
    770                 775                 780
Met Phe Asp Gln Arg Val Ile Ala Tyr Phe Gly Asn Pro Lys Glu Ile
785                 790                 795                 800
Ile Ala Asp Asn Asp His Ile Phe Thr Ser Gln Thr Trp Lys Asp Phe
                805                 810                 815
Ala His Lys Tyr Asn Phe Val Met Lys Phe Ser Leu Pro Tyr Arg Pro
            820                 825                 830
Gln Thr Asp Gly Gln Thr Glu Arg Thr Asn Gln Thr Val Glu Lys Leu
        835                 840                 845
Leu Arg Cys Val Cys Ser Thr His Pro Asn Thr Trp Val Asp His Ile
    850                 855                 860
Ser Leu Val Gln Gln Ser Tyr Asn Asn Ala Ile His Ser Ala Thr Gln
865                 870                 875                 880
Met Thr Pro Phe Glu Ile Val His Arg Tyr Ser Pro Ala Leu Ser Pro
                885                 890                 895
Leu Glu Leu Pro Ser Phe Ser Asp Lys Thr Asp Glu Asn Ser Gln Glu
            900                 905                 910
Thr Ile Gln Val Phe Gln Thr Val Lys Glu His Leu Asn Thr Asn Asn
        915                 920                 925
Ile Lys Met Lys Lys Tyr Phe Asp Met Lys Ile Gln Glu Ile Glu Glu
    930                 935                 940
Phe Gln Pro Gly Asp Leu Val Met Val Lys Arg Thr Lys Thr Ala Phe
945                 950                 955                 960
Leu Tyr Thr Asn Asn Arg Gln Thr Glu Ser Gln Ile Met Ser Glu Leu
                965                 970                 975
Pro Phe Thr Ile Ala Ser Lys Arg Ile Lys Tyr Leu Gly Ile Gln Leu
            980                 985                 990
Thr Arg Glu Val Lys Asp Leu Phe Lys Glu Asn Tyr Lys
        995                 1000                1005
```

<210> 93
<211> 195
<212> PRT
<213> Homo sapiens

<400> 93
Gly Pro Arg Leu Ala His Gly Thr Thr Thr Leu Ala Phe Arg Phe Arg
1 5 10 15
His Gly Val Ile Ala Ala Ala Asp Thr Arg Ser Ser Cys Gly Ser Tyr
20 25 30
Val Ala Cys Pro Ala Ser Cys Lys Val Ile Pro Val His Gln His Leu
35 40 45
Leu Gly Thr Thr Ser Gly Thr Ser Ala Asp Cys Ala Thr Trp Tyr Arg
50 55 60
Val Leu Gln Arg Glu Leu Arg Leu Arg Glu Leu Arg Glu Gly Gln Leu
65 70 75 80
Pro Ser Val Ala Ser Ala Ala Lys Leu Leu Ser Ala Met Met Ser Gln
85 90 95
Tyr Arg Gly Leu Asp Leu Cys Val Ala Thr Ala Leu Cys Gly Trp Asp
100 105 110
Arg Ser Gly Pro Glu Leu Phe Tyr Val Tyr Ser Asp Gly Thr Arg Leu
115 120 125
Gln Gly Asp Ile Phe Ser Val Gly Ser Gly Ser Pro Tyr Ala Tyr Gly
130 135 140
Val Leu Asp Arg Gly Tyr Arg Tyr Asp Met Ser Thr Gln Glu Ala Tyr
145 150 155 160
Ala Leu Ala Arg Cys Ala Val Ala His Ala Thr His Arg Asp Ala Tyr
165 170 175
Ser Gly Gly Ser Val Asp Leu Phe His Val Arg Glu Ser Gly Trp Glu
180 185 190
His Val Ser
195

<210> 94
<211> 198
<212> PRT
<213> Homo sapiens

<400> 94
Ser Ile Met Ser Tyr Asn Gly Gly Ala Ile Met Ala Met Lys Gly Lys
1 5 10 15
Asn Arg Val Ala Ile Ala Ala Asp Arg His Phe Gly Ile Gln Ala Gln
20 25 30
Met Val Thr Thr Asp Phe Gln Glu Ile Phe Pro Met Gly Gly Trp Leu
35 40 45
Tyr Ile Gly Leu Ala Gly Leu Ala Thr Asp Val Gln Arg Val Ala Gln
50 55 60
Cys Leu Lys Phe Gln Leu Asn Leu Tyr Glu Leu Lys Glu Gly Gln Gln
65 70 75 80
Ile Lys Pro Tyr Thr Phe Thr Ser Met Val Ala Asn Phe Leu Tyr Glu
85 90 95
Lys His Phe Gly Pro Tyr Tyr Thr Asp Pro Val Ile Ala Gly Leu Asp
100 105 110
Leu Lys Thr Phe Lys Pro Phe Ser Cys Ser Leu Asp Leu Ile Gly Phe
115 120 125
Pro Met Val Thr Asp Asp Phe Val Val Asn Gly Ser Tyr Ala Glu Gln
130 135 140
Met Tyr Gly Met Cys Glu Ser Leu Trp Glu Pro Asn Met Asp Pro Glu
145 150 155 160
His Pro Phe Glu Thr Ile Ser Pro Ala Met Leu Asn Ala Val Asp Trp
165 170 175

Gly Ala Gly Ser Gly Met Gly Val Ile Ile His Ile Thr Lys Lys Asp
180 185 190
Lys Ile Thr Thr Arg Thr
195